

Making arms control stick?

It is a good sign that the superpowers have fallen silent, in public, about the intricacies of the Geneva negotiations, which suggests that they are getting somewhere. But what will happen next?

Two months have now passed since the United States appears successfully to have persuaded its European allies that they had better swallow their misgivings and agree that Europe would be a safer place without missiles of intermediate range. The United States had an easy case to make: Europe seemed content enough with the strategic balance of the mid-1970s, before Soviet SS-20 missiles began appearing in the East, so what could be wrong with returning to that state of affairs? West Germany's reluctance, born of the view that Central Europe was even then a dangerous place, was natural enough, but has been turned, at least for the time being. Britain and France (an ally if not technically a member of the North Atlantic Treaty Organization) were more concerned that their independent nuclear forces should survive an agreement on missiles of intermediate range, and appear to have been satisfied, again for the time being. But these half-reluctant partners in what seems to be an impending arms control agreement, not to mention the fellow-allies such as the Netherlands and Belgium, which will no doubt be heartily glad to see the back of the US Pershing missiles, had better be thinking now of what will happen next.

First, all concerned (in the Soviet Union as well) had better acknowledge that circumstances have much changed since the mid-1970s, when what was called *détente* was at its peak. For what it is worth, it has never been clear in the West what went so sour then. It is true that Mr Richard Nixon, the architect (with Dr Henry Kissinger) of the accommodation between East and West, had lost his job as president of the United States, but that can have made very little difference to the understanding that had been reached. And while the growing disenchantment of the United States with the war in Vietnam may have seemed to those then in charge in the Soviet Union to be a sign of incipient weakness, surely that knowledge would also have worked in favour of keeping the understanding alive. Yet the SS-20 missiles began to make their appearance in the East without formal announcement. The decade, it will be recalled, ended with the considerable Soviet presence in Afghanistan, still continuing. It would have been easier for those now worried about the damage that even short-range nuclear weapons could cause in Central Europe to accept the impending deal on intermediate missiles if there were a better understanding of why the earlier climate of *détente* had melted away.

By extension, it follows that the agreement the two superpowers appear likely to reach later in the year will be the more durable, even palatable, if it is accompanied by some kind of restoration of the climate of the mid-1970s. The obvious difficulty is that, while President Reagan may be well placed, in the closing year or so of his presidency, to make (and then carry through the US Senate) an agreement to remove intermediate missiles from Europe, it is much less likely that he could reach a broader political settlement at this stage. Mr Mikhail Gorbachev may therefore have to wait in patience for the broader understanding he seems to want until after the elections next November. Would it not, in the interval, be sensible to try to breathe new life into the Helsinki agreements, already negotiated and signed by all concerned?

Meanwhile, on the assumption that there will be an agree-

ment on intermediate missiles before the year is out, it is prudent to anticipate, first, that the agreement will itself engender some kind of rekindling of *détente* and, second, that the good effects will begin to fade unless there is the prospect of another agreement in sight. The obvious difficulty there is that the Geneva talks have shown that the obviously necessary agreement on strategic arms will not be possible without the containment of military technology in space, likely to be unacceptable at least for the duration of this presidency. This is why the time has now come to dust off the comprehensive test-ban treaty, even as amended by some of last year's ingenious schemes for allowing a quota of exceptions on either side. Both partners in the Geneva negotiations deserve congratulations for having brought things this far, but their efforts will have been in vain if they do not arrange to maintain the momentum of agreement. □

New brooms for old

The new British government's plans offer no fresh hope for research, but some opportunities.

PREDICTABLY, the newly re-elected British government has embarked with zeal on its new legislative programme (see p.6). If the prospect of being hanged clarifies the mind, reprieve (even from the passing threat of electoral defeat) is a powerful invigorant. Yet nothing in what the government advertised in the Queen's speech last week promises a respite from hardship for Britain's research community. Keeping "firm control of public expenditure", so as to keep inflation in check and further to reduce the burden of British taxes, remains the objective. Nowhere does the government acknowledge that the time may have come for adding to its economics of good housekeeping a strategy for research-led innovation.

Moreover, the promise of legislative frenzy notwithstanding, many of the uncertainties that afflict research persist. The Department of Education and Science hopes that its education bill (fondly known by civil servants as "Gerbill", for Great Educational Reform Bill, with scant regard for the principle that puns founder on poor spelling) will be at the top of the legislative queue, yet among the plans for changing the rules on higher education canvassed over the past few years, only that for taking polytechnics away from local education authority control rated a mention last week. The scheme for financing universities by means of contracts for the "provision of educational services" will no doubt be found in Gerbill when that is published in the autumn; between now and then, there is a chance that the government's views may be changeable by persuasion. It is not even clear whether the government will move to abolish academic tenure in this long parliamentary session, which again suggests that there may still be room for manoeuvre. Sir Keith Joseph's mean-minded plan for extinguishing tenure by means of a network of parliamentary commissions is an administrative nightmare the new administration cannot welcome, yet pride will impel it to do something. The best hope is that it will agree that the issue of tenure should be subsumed in a wider examination of the conditions on which academics are employed, and



which is in any case necessary to make the British system flexible.

As always in British education bills, the big issues will concern the schools. Britain is not the only industrial nation whose secondary education is no longer matched to modern needs, so that some attention to these questions is entirely proper. The British government confirmed last week its sensible plan (see *Nature* 327, 448; 1987) to institute a nationally applicable curriculum for secondary schools, although it is far from clear what that curriculum will consist of, and whether it will suit the need. But the government seems also determined to muddy the issue by its proposals for weakening the links between schools and local education authorities (which may be desirable) and for encouraging intending secondary-school students to apply for places at schools of their choice (which, on the face of things, belies the purpose of a national curriculum — a degree of uniformity). Both proposals will be contentious. The second, justified by the rubric "parental choice", is inadequately thought through and may yet prove to be what its critics say it is intended to be, a device for allowing educational ghettos to persist in British public education. The Secretary of State for Education and Science, Mr Kenneth Baker, will have to think quickly to find a means of avoiding this pitfall before the autumn. □

Centres for competition

The US National Science Foundation is embarking on overdue growth — not without risk.

It is no accident that the US National Science Foundation (NSF) has become the most favoured of the Reagan administration's instruments for research support. Its chief claim on public affection is the diversity of its projects, almost all of which are managed by university researchers (as with simple project grants) or by consortia of universities (running observatories and other central services). Although NSF does from time to time spend substantial sums of money (on telescopes, for example) its bills are always small compared with those of, say, the Department of Energy, traditionally responsible for building particle accelerators, or the National Aeronautics and Space Administration, which has spent \$1,200 million on the Hubble Space Telescope. Moreover, NSF differs from other agencies spending money on research, such as the National Institutes of Health (NIH), in its ability and even readiness to turn its mind to what appear to be national problems. Now that the United States has taken fright that US industry may no longer be internationally competitive, NSF has manfully stepped forward with an offer to help. No wonder that the administration plans that NSF's budget should be doubled over the next five years. (It should be clear in a few weeks whether Congress will agree to the first tranche of 17 per cent.)

But how can a grant-making organization which must almost passively respond to the requests that others make upon it safely direct a research community's efforts towards industrially, and even commercially, strategic goals? The most important of NSF's new schemes is that for founding a network of centres for strategic research in well-specified areas of endeavour. The idea, very much that of Dr Erich Bloch, NSF's director, is that the centres should be linked geographically and otherwise with universities but that they should be managed separately, and that potential industrial beneficiaries should be involved. Wisely, NSF has taken outside advice before steaming ahead. Whether NSF will be grateful for the less than full-throated endorsement of its plans that has emerged (see p. 5) is another matter.

Caution is nevertheless appropriate. It is still within living memory (in 1970) that NSF, under an earlier management, embarked for similar reasons on its ill-starred programme called "Research Applied to National Needs" (RANN). More recently, at the beginning of the first Reagan administration, the US

Department of Commerce was flirting with a scheme not very different from that now being canvassed by NSF, but of which very little came. The essence of the difficulty, especially acute in an economy at least structurally as competitive as that of the United States, is that of arranging that people who are not directly engaged as cogs in the competitive machine — academics, NSF officials and even industrialists in their role as public servants — should accurately tell what projects will best ensure the rapid improvement of industrial technology. There is plenty of experience elsewhere, in Britain for example, to show that people asked to pick winners in circumstances like these usually choose well for races already run.

Luckily, for NSF, there are many fields in which well-found research centres could have a rapid and beneficial influence in US industry. The great fuss in recent weeks about superconductivity in ternary oxides with perovskite structure, for example, points to a need for means of working out atomic structures in circumstances very different from those in which classical X-ray diffraction is conventionally applicable — where disorder from one unit cell to the next may be crucial, for example. Exactly the same need arises in connection with one of last year's wonders of solid-state physics, the puzzle of quasicrystalline structures with fivefold symmetry — a problem fast running into the sand for want of data. But semiconductor surfaces and solid structures designed to merge one lattice smoothly into another (GaAs is the favourite) predicate the same need. The advantage of working in such a field is that there is a crying need for basic understanding as well as for the practical techniques that will ultimately benefit industry. The danger, which NSF's advisory panel at the US National Academy of Science might more clearly have spelled out, is that attempts to plug gaps already opened between the United States and Japan are likely to lead nowhere, however appealing they may be to the US Congress.

That is why NSF will need strong nerves in spending its enlarged budget. For much of its 30 years, the agency has been small and relatively inconspicuous, able even to make mistakes without bringing trouble down upon itself. As it grows, the pressure to respond to popular causes will also increase and, with it, the temptation to make mistakes. Much will hang on the plan for science and technology centres now being hatched. □

Who goes where?

Estimates of the loss by emigration of skilled people are not exact.

THE Royal Society's attempt to tell how rapid is the loss of technically trained people from the United Kingdom, published this week (see page 27), is a good illustration of the difficulty of measuring the effects of the migration, commonly known as the "brain-drain". Exactly similar difficulties have been encountered in earlier and less well-conducted surveys — in Britain in 1968, for example — and would almost certainly arise in places where the loss of skilled people is at once greater and more serious in its consequences, in India for example. For one thing, the statistics are always out of date (the latest British study stops short with figures for 1985). For another, but more important, there is no way of telling which of those who leave for appointments overseas will never return; accurate counts would be possible only after the passage of a lifespan.

But, whatever the statistics, their implications for public policy are not easily discerned. There is no basis for asking technically trained people out of patriotism to stay at home, especially when most of those who emigrate do so, in the first instance anyway, so as to improve their skills. And, in any case, there is, in the West, no legal basis for restraining them. But this week's study almost certainly underestimates the rate at which able people are now leaving Britain, even if the chief consequence of the recent decline of the research enterprise is probably that fewer able people are being produced. □

French AIDS campaign launched at long last

- No compulsory screening but free tests on tap
- Campaign with 24 million brochures

Paris

A LONG-AWAITED national campaign to combat the spread of AIDS (acquired immune deficiency syndrome) in France was outlined by health minister Michèle Barzach at a cabinet meeting on 24 June. With 1,632 cases of AIDS and an estimated 150,000–200,000 carriers of antibodies to the human immunodeficiency virus (HIV), France is top of the league in Europe and in fourth place among industrialized countries, yet has been conspicuously slow to announce national countermeasures. This is particularly ironic given the international kudos attracted by the Pasteur Institute for its early success in isolating the AIDS virus.

Barzach, a general practitioner, has based her proposals on a report, (*The Hospital and AIDS*), commissioned by her ministry last December and published at the beginning of June, from a multi-disciplinary working party headed by social services minister Jean Choussat. The four-point national campaign embraces prevention, care, research and international cooperation.

The government's proposals also follow the working party's recommendations in emphasizing that there will be no programme of compulsory screening for HIV antibodies. This provision has already attracted the ire of the extreme right-wing National Front party, whose AIDS policy, published the same day as Barzach's, advocates a mandatory twice-yearly screening of all French citizens at an estimated cost of FF16,000–25,000 million (£1,600 million) a year, the compulsory follow-up of all HIV-positive people and the institutionalization of confirmed AIDS cases.

Rather than a national screening campaign, which, it is argued, would be prohibitively expensive as well as ineffective, Barzach plans to make free, voluntary and anonymous screening widely available at special centres throughout France. There will also be a concerted information campaign, directed especially at young people and health workers, which will be launched in the cinema, on television, through posters and with 24 million brochures distributed to everyone on the electoral list. The cost of the brochures, estimated at FF20 million, will be borne by the post office.

Other preventive measures include the systematic, anonymous screening of blood whenever a sample is taken from a

patient, the supply of syringes without prescription and freedom for manufacturers to advertise condoms.

With the search for an AIDS vaccine declared a cause "of national importance" to which the public is invited to donate, Barzach intends to provide an extra FF100 million (£10 million) for research into new methods of treatment and diagnosis and the development of a vaccine. Meanwhile, eleven centres will be opened in October in Paris and other French cities for the care and monitoring of AIDS patients. In addition, it is intended that general hospitals should create small AIDS units, with up to 20 beds and with an emphasis on day-care where possible.

With no official government strategy to combat the spread of AIDS until this week, one department (county) had already drawn up its own plan. The Alpes-Maritime region of southern France, where both AIDS and National Front membership are relatively widespread, a phenomenon attributed to a high immigrant population and widespread drug abuse, intends to make screening for HIV antibodies mandatory and to create a register of carriers of the AIDS virus, ostensibly to permit effective epidemiological analysis.

This move, which could contravene civil rights laws, has been condemned by the health ministry. Alpes-Maritime officials, who had hoped to rely on a proviso in the constitution that authorizes local action to control the spread of venereal diseases, have been told by Barzach that AIDS does not come under this category. The minister is also adamant that a coordinated national policy, not a series of local initiatives, is what is required.

Prevention is likely to be more difficult in France than in some other countries. With a predominantly Catholic population, France has an uphill struggle ahead to make the use of condoms widespread. The director of one leading French manufacturer of surgical goods has estimated that only 17 per cent of families practising some form of contraception choose the condom, compared with over 70 per cent in Japan, which has one of the lowest rates of AIDS cases in the world.

The French episcopacy has recently found itself in an extraordinary double-bind. As a prophylactic against AIDS, use of the condom is not discouraged by the Church, but as a contraceptive it remains proscribed.

Peter Coles

NASA tests too hasty?

Washington

A SPECIAL panel of the National Research Council evaluating the redesign of the space shuttle's booster rockets last week issued an interim report critical of NASA (National Aeronautics and Space Administration) for failing to pursue alternatives for certain critical aspects of the design "with sufficient vigour". By its neglect, the panel warns, NASA is risking even greater delays in its timetable.

One particular concern is the procedure for testing the new design for the booster joints. Because several test firings will not be feasible, the joints will be tested by intentionally introducing flaws to see if they fail as in the Challenger accident. The panel is worried that the intended flaws may reveal problems for which NASA has no solutions.

A full-scale full-duration test firing of the newly designed joints is planned for 23 August. J.P.

This week's prizes

London

THE architect of India's Green Revolution, Dr M. S. Swaminathan, is the first recipient of a new \$200,000 General Foods World Food Prize for his contribution to improving the world's food supply. Swaminathan, director general of the International Rice Research Institute in the Philippines, has worked to improve production of rice and wheat in India and Pakistan.

The \$300,000 Kyoto Prize for Basic Sciences has been awarded to Jan Oort, a Dutch astronomer who described the phenomenon of spinning galaxies 60 years ago. The prize honours his outstanding contribution to the deeper understanding of the universe. K.J.

University for Tanzania

London

THE Tanzanian Party leader, Mwalimu Nyerere, has asked the Soviet Union to help establish a university of chemistry and technology at Mbeya in western Tanzania. A college of technology is already being built there, with Soviet assistance, but now the Tanzanians feel that it would be "even better" to convert it into a full-scale university. V.R.

Plant skills lacking

London

A SHORTAGE of skilled scientists at the postdoctoral level could seriously hamper the efforts of British companies to compete in the field of biotechnology. So says a report from the Science and Engineering Research Council, which cites plant molecular technology and bioprocess technology as areas of particular shortage. The report is discussed on page 96 of this issue. C.W.

Joint survey by Japan and France in South Pacific

Tokyo

JAPAN's Science and Technology Agency and France's IFREMER last week signed an agreement to carry out a joint survey of back-arc basins in the South Pacific using deep-sea submersibles from both countries.

The multi-million dollar project, named STARMER-1, was first proposed last year by Japan and France at a meeting in the Cook Islands of the Committee for Coordination of Joint Prospecting for Mineral Resources in South Pacific Off-shore Areas and was warmly welcomed by South Pacific island nations, according to Dr Honza of the Geological Survey of Japan, one of the project scientists. Expected to run for five years, the survey will be joined by scientists from the South Pacific islands, as well as France, Japan and the United States, and comes at a time when Japan, at the urging of the United States, is showing increased interest in this region.

The *Kaiyo*, a twin-hulled research vessel belonging to the Japan Marine Science and Technology Center of the Science and Technology Agency, will set sail in November to carry out a survey of the North Fiji Basin. The basin lies between the colliding Pacific and Indian plates and has a complex network of spreading ridges on which may be located hydrothermal vents, vent organisms and mineral deposits.

The *Kaiyo*'s multibeam side-scan sonar (SEABEAM) will map the sea floor of the basin, which ranges from 2,000 to 4,000 m in depth, and hydrographic equipment and a deep-tow vehicle will be used to try to locate hydrothermal vents and suitable sites for the French submersible *Nautille* to dive to in late 1988 and/or early 1989.

At a later stage, Japan will bring the submersibles *Shinkai 2000* (depth range 2,000 m) and *Shinkai 6000* to the survey area — the latter submersible is due to be built by 1988 and will have a depth range of 6,500 m, similar to *Nautille* (6,000 m). Japan's recently completed *Dolphin 3K*, an unmanned vehicle equipped with manipulator arms, television and CTD (conductivity-temperature-depth meter), will also be let loose to scan and sample the sea floor while tethered to the ship by several kilometres of optical-fibre cable.

Apart from locating vents and associated fauna and mineral deposits, STARMER-1 will aim at elucidating the mechanisms of seafloor spreading in back-arc basins and will also provide an opportunity to study the microplates in the buffer zone between the Indian and Pacific plates, according to Honza.

Marine scientists affiliated with the

Ministry of International Trade and Industry, the Ministry of Education, Science and Culture, and the Environmental Agency, as well as the Science and Technology Agency, will take part in the project. But funding, to be split equally



Japanese deep-sea submersible.

between Japan and France, will come entirely from the Science and Technology Agency on the Japanese side, and the agency expects to use Y200–300 million (about \$2 million) out of its special promotion funds for this fiscal year to cover the costs of the *Kaiyo* cruise in 1987–88, according to Masato Chijiya, director of the Ocean Development Division of the agency.

The Ministry of Education, quick to spot an opportunity to save money, has proposed another joint survey with France in 1988, also using the *Nautille*, as a sequel to the KAIKO project (see *Nature* 308, 102; 1984). Although independent of STARMER-1 — the proposed survey area of KAIKO-2 is the Nankai Trough off the Pacific coast of Japan — the presence of the *Nautille* and its mother ship the *Nadir* in the Pacific in 1988–89 would greatly reduce the costs of KAIKO-2.

David Swinbanks

India says Union Carbide must pay compensation for Bhopal

New Delhi

INDIAN officials say that Union Carbide cannot avoid paying hefty compensation to victims of the Bhopal gas tragedy by trying to prove that the accident was caused by sabotage. Sources said that the US company is liable to pay irrespective of how the accident occurred.

The officials were reacting to a report in the *New York Times* on 23 June saying that Union Carbide has found "new witnesses, documents and scientific evidence" proving that the 1984 leak of toxic methyl isocyanate (MIC) gas was caused by sabotage by a company employee. They said the sabotage theory is untenable and alleged that Carbide might be fabricating evidence. "In any case", said one official, "it is good for us that Carbide is coming out with everything they have. It will help our lawyers."

India is suing Union Carbide for \$3,300 million in damages after having rejected an out of court settlement offer of \$300–350 million. The lawsuit over the Bhopal disaster, which killed about 2,000 and injured another 200,000, has been bogged down in the Indian courts since September 1986 when India first filed the claim.

Evidence to prove sabotage was filed by Carbide through a 160-page affidavit in the Bhopal court six months ago. According to the *New York Times*, the company, which is nearing the end of its 16-month investigation, has obtained new documents and witnesses including an operator who allegedly committed the sabotage. According to Carbide, the operator removed a pressure gauge connected to the MIC tank and connected a water hose. Carbide claims to have a witness who

noticed the water hose and the missing pressure gauge. Other proof includes altered log books to erase evidence that addition of water was deliberate.

Carbide's legal strategy has been to avoid paying several million dollars of compensation by showing that the accident was caused by deliberate action by an employee. An Indian technical team that investigated the accident had agreed that entry of water into the MIC tank triggered the accident. But it made clear that bad design of the plant, wrong choice of materials for its construction, and inadequate provision of control and alarm systems were chiefly responsible for the accident running out of control. "We do not believe in the sabotage theory because it is impossible for a low-level operator to be clever enough to know the consequences of his act which we ourselves could determine only after weeks of analysis", said a scientist connected with the investigation.

Indian officials say they are not worried about what theory is advanced by Carbide as far as the lawsuit goes. "All we say", said an official, "is that the plant was so badly designed that conditions for the accident were inherent and extant." He likened the Carbide-managed Bhopal plant to a room with a box of explosives at a corner, a detonator in another corner, wires in between and a roof full of holes. "All conditions existed for the room to explode with the fall of a big stone over the detonator from the hole in the roof", he said. "Whether the stone was dropped accidentally or deliberately does not make any difference to our case."

K.S. Jayaraman

Planned NSF centres to boost economy may gain from funding

Washington

A PANEL of the National Academy of Sciences has issued an enthusiastic, but at the same time cautious, endorsement of the plans of the National Science Foundation (NSF) to establish science and technology centres at US universities*. Acknowledging that these centres can make "significant contributions to science and the nation's economic competitiveness", the panel also warns that great care will be needed to keep the programme "in proper balance" with other ways of supporting US science.

President Ronald Reagan proposed that federal agencies establish science and technology centres in his 1987 State of the Union address as a way to improve US economic competitiveness (see *Nature* 325, 473; 1987). Although other agencies have also embarked on efforts to found such centres, the academy's report is a direct response to a request by NSF director Erich Bloch for help in formulating his own agency's plans.

The panel, chaired by Stanford University professor Richard Zare, urged NSF to let the scientific community determine the goals for the centres. By suggesting particular topics, the panel worried that NSF might make researchers "inadvertently steer away from areas of even greater promise".

The academy panel is also anxious that the science and technology centres should not divert funds from individual research support. The worst case, says Zare, would be for Congress to decree that the centres should be created without adding the money to NSF's budget to pay for them.

NSF is alert to the need to balance its research support. The agency has said it plans to spend no more than 10 per cent of its budget on the centres, including those already established. Alan Leshner, director of the newly formed office that will coordinate development of NSF's centres, says that other programmes will not be raided for the funds needed to start them. Leshner says the 1988 budget request for the science and technology centres is about \$30 million, "but if we get less, we will do less".

Zare's panel also argues that making the fruits of a centre's research available to industry will require "outreach" programmes such as seminars and conferences. Knowledge transfer is a "body contact sport", says Zare. He specifically suggests that industry should not be required to contribute financially to the centres, participating instead as col-

leagues. But improved conditions for conducting scientific research provide only part of the solution of the problem of US competitiveness internationally. Zare also says that economic policy and labour reform will also have to be part of a strategy to improve US economic strength.

The panel suggests several models for the organization of science and technology



NSF director Erich Bloch

centres. While a central intellectual theme could justify the creation of a centre to bring together a critical mass of research talent, on other occasions a large piece of research equipment might form the focus of a centre, but there could also be "centres without walls" where interaction took place by means of electronic mail. The intellectual home of each centre should be a university campus, although a physical plant could be off-campus.

The panel suggests that proposals for centres should be judged for their scientific value by peer review, and that final decisions should be based on the recommendations of a multidisciplinary panel. Research awards would be for up to nine years, with no centre receiving fewer than six years of financial support. Annual NSF investment would be between \$500,000 and \$10 million.

The panel also says there is a risk that centres may become "institutionalized", losing some of the flexibility Zare's panel hopes for them. Its report also warns of the dangers of become too enamoured of the concept of interdisciplinary research frequently advocated among the reasons for founding science and technology centres. It would be "unfortunate" if good projects were not supported "simply because they are not regarded as sufficiently cross-disciplinary".

Leshner says NSF will implement many of the academy's recommendations, although there will be differences. Awards may not be for nine years, and there may be a financial component to industry involvement. Leshner hopes that a formal announcement of the programme

can be made after the National Science Board meets again in August.

In the long run, the centres may provide a new political constituency for basic research. Just as medical research is justified by reference to the health of the United States and military research by reference to national security, Zare believes economic imperatives may become a battle-cry for putting more money into basic research.

Joseph Palca

• Prospects that money will be found for the new science and technology centres brightened last week as good progress was made in Congress on authorization bills (which describe how funds should be spent) and appropriation bills (which provide the funds) for the National Science Foundation. The House authorization bill was passed and provided for the same 17 per cent increase in the NSF budget that had been requested by President Reagan. The more critical House appropriations bill hit some rough spots at the committee stage and at one point seemed likely to shed \$100 million. But action at the last moment restored most of the cuts. The final vote before the House will take place this week and opposition to the NSF budget increase remains possible. When the Senate considers the appropriations bill, Senator William Proxmire (Democrat, Wisconsin) remains a figure to be reckoned with.

Alun Anderson

Chinese rules for study abroad

London

CHINA'S State Education Commission last month published its "interim provisions" for sending personnel to study abroad. These provisions had in fact been in operation for more than six months, and had already been discussed to some extent in the Chinese media. Their publication in full suggests that they have not so far been implemented with the zeal the Education Commission had expected.

The "interim provisions" put special emphasis on the contractual basis of studying abroad. Study trips are to be predominantly in the applied sciences, and personnel sent abroad under government auspices (including the local government level) must sign a contract with the organization sending them, specifying the objectives and subject of the visit, and the minimum length of time during which the student will work for the sending organization after the trip. The political qualifications of those sent abroad are stressed; they must "ardently love the motherland and socialism, have good ideological and moral qualities, have distinguished themselves in practical work and study, and have served socialist modernization actively".

Vera Rich

* *Science and Technology Centers: Principles and Guidelines*, National Academy of Sciences, Washington, DC, 1987.

New space agency proposed by West German parliament

Munich

WEST Germany will coordinate its space programme through a new private agency, if the proposal of Research and Technology Minister Heinz Riesenhuber is accepted by the German Bundestag (parliament). The proposal, likely to be approved by the Bundestag in September, was introduced on 24 June.

In an unrelated move, the Bundestag's budget committee agreed to double the 1987 West German space budget to DM 502 million. The money is divided among three quite costly European Space Agency (ESA) projects — the booster rocket Ariane, the space shuttle Hermes and the manned space station Columbus, which is to be built by the United States with ESA cooperation. The total cost to West Germany for the three is now estimated at about DM 8,000 million.

Until now, the West German space programme has been administered by the German Aerospace Research Establishment (DFVLR), which in turn reports to the Research and Technology Ministry (BMFT). This led to an uncomfortable situation, as the DFVLR was politically responsible for assigning contracts as well as for carrying them out.

The proposed agency, the German Agency for Space Flight (DARA), would negotiate with ESA and other space authorities on West Germany's behalf and coordinate industry and scientists as well as ministries that have an interest in the

space programme. Its budget would probably be several tens of millions of deutschmarks. The DFVLR would give up its political role and simply carry out the contracts awarded it by the agency.

DARA, if indeed approved, will be a private agency, which is not bound by West German law to pay salaries on the public service wage scale as are government institutions. The BMFT hopes that the higher pay will attract better people to the agency.

Not surprisingly, the DFVLR is not happy with the Riesenhuber plan, which would relieve it of up to 100 employees in the coming years and reduce its scope by about 20 per cent. The chairman of the DFVLR, Walter Kröll, has submitted a reorganization plan of his own under which political and scientific functions of the space programme would also be divided.

The biggest advantage of the Kröll plan, said DFVLR official Rüdiger Clausen, is

that it does not require a larger investment. Founding an entirely new agency, especially if it is far from the Cologne headquarters of the DFVLR, will be very costly. The location of DARA has already become a hot political topic in various *Länder*.

But the most logical place to put the new agency would be in Cologne, possibly in the same building as the DFVLR, hinted Bundestag member Catenhusen. This plan would foil the apparent designs of Bavarian Minister-President Franz-Josef Strauss to bring the entire German space programme to Bavaria. Earlier this year, Munich (Bavaria's capital) tried to outbid Cologne for the right to expand its DFVLR facilities. The attempt failed, as Cologne raised its contribution as well. The facilities remain split between the two cities.

Even though a positive verdict is expected for DARA, the larger question of German participation in the Ariane, Hermes and Columbus projects is still wide open. Catenhusen and others vocally oppose continued participation, at least in Hermes and Columbus, as the stakes inevitably go up. **Steven Dickman**

Education reform is priority of the British government

London

THE British government last week unveiled its programme for the parliamentary session to November next year which, among other things, will substantially change the education system and patent law. The programme, outlined by the Queen at the state opening of parliament, is one of the heaviest for some time.

The new education legislation, one of 17 bills to be processed by the government, will ensure that parents are given more choice in the style and location of their children's education and there will be a new curriculum for schools, subjecting pupils to four formal achievement tests by the age of 16.

Despite the government's overall majority of just over 100 seats in the House of Commons, the proposed legislation is likely to provoke much hostility from those who claim that the deficiencies of the current system are the result of inadequate funding and cannot be cured by the legislation.

Certainly, studies in the past few years have highlighted the shortage of science teachers, particularly in physics; this shortfall and others in mathematics and the applied sciences/technology are due to an inability to match the salaries offered by industry to technical and numerate graduates.

The legislation is also meant to guarantee that the nation's education system is

controlled more centrally through the Department of Education and Science (DES) which, under the current system, has less influence on schooling in some areas than the local education authority. In further education, the same philosophy applies, and the new Education Act will allow the 29 polytechnics and other colleges to become "free-standing corporate bodies outside local authority control". It will also ensure that the curriculum adopted by these colleges reflects national needs and that "their governing bodies would have strong representation from industry, commerce and the professions".

The Queen's speech also confirmed the belief (see *Nature* 327, 549; 1987) that the government would reintroduce legislation, jettisoned because of time constraints before the general election, to change copyright and patent law.

Apart from devising a modern legislative framework to protect British products from piracy, the current gaps in copyright law, dealing with computer outputs and databases, are to be plugged. The legislation will also seek to provide better protection for 'intellectual property'. "The main provisions are to encourage investment in new design by introducing or extending protection to designs; reform patent law, among other things, to increase competition in patent advice and reduce the expense of patent litigation."

Bill Johnstone



Transgenic sow

London

A TRANSGENIC sow and her litter, produced as part of a collaborative research programme between Animal Biotechnology Cambridge Ltd (ABC) and Embryogen Inc., of Ohio. ABC is a recently formed company that has taken over the former government-funded Animal Research Station at Cambridge, England. Among ABC's first venture partners is a consortium of city investors, including Biotechnology Investments Ltd, which has committed more than £1 million to the new company. Under the scientific directorship of Dr Christopher Polge, ABC's first commercial project is a three-year programme to develop techniques for large-scale production and freezing of cattle embryos suitable for transplant. Other areas of research include transgenesis and sex pre-determination.

Simon Hadlington

Agreement nears on protection of the Earth's ozone layer

Brussels

KEY actors in the continuing drama surrounding an international protocol to protect the Earth's ozone layer by regulating chlorofluorocarbons (CFCs) are meeting in Brussels this week for final discussions and negotiations before governments are asked to sign the far-reaching agreement in September.

The meeting, an informal gathering of representatives of signatories of the Vienna Convention for the Protection of the Ozone Layer and its chairman, United Nations Environment Programme director Mostafa Tolba, has been aimed at settling issues left outstanding after the last round of negotiations at Geneva in April. But instead of the entourages of advisers, government representatives, non-government organizations, scientists and economists familiar at previous Vienna Convention meetings, this week's discussions have involved only the heads of delegations of some signatories, which marks a shift of the negotiations from the technical level to the political.

One of the crucial issues for this week's discussions has been the size of the reduction in global production and consumption of CFCs, which are used as aerosol propellants, foam-blowing agents, refrigerants and solvents. The draft agreement in April called for a freeze of CFCs 11, 12 and 113 at the 1986 level, a 20 per cent cutback two years later and a second stage reduction of a further 30 per cent, either after six or eight years.

At the start of this week's meeting, there were signs of some backtracking on that position, especially by the European Community. A commission representative said in Brussels last week that a simple 20 per cent reduction "represents what is necessary and achievable in the short term". While the official US position is that CFCs should eventually be eliminated, that position is opposed by officials in the Interior Department and the Commerce Department, who last week put their arguments to President Reagan.

Negotiations in Brussels this week have also dealt with the question of trade in CFCs, including provisions for regulating and monitoring trade between producer and non-producer countries.

The scientific evidence that first led to concern about CFCs is being discussed by an international committee set up by the US National Aeronautics and Space Administration (NASA), the Ozone Trends Panel, which met last week. The panel and its working groups comprise more than 100 people concerned with all aspects of research into stratospheric ozone depletion, both total column and

vertical profile, as well as the role of CFCs in ozone destruction. The panel is also reviewing satellite and ground-based ozone data, partly as a response to evidence submitted to the US Congress earlier this year by NASA scientist Don Heath which concluded, on the basis of satellite data, that global ozone depletion has been far more severe than originally predicted.

Members of the working group on the Antarctic ozone hole are awaiting the result of this year's Antarctic Ozone Expedition, expected to provide more information on whether the ozone hole is

caused by chemicals such as CFCs or is a dynamic process. New data, not yet published, are consistent with the chlorine hypothesis.

The panel will also study a decrease in global ozone in the early 1960s to tell whether extra nitrogen produced by nuclear testing could have been responsible.

NASA scientist Dr Robert Watson, chairman of the Ozone Trends Panel, says its report has "become a monumental work" and is unlikely to be completed before the scheduled September meeting for the signing of an international protocol on CFCs. "But the politicians have enough to go on now", Watson notes. "Anything we find will only make the situation worse for CFCs, not better."

Kathy Johnston

University proposed to span British Commonwealth

London

AMBITIOUS proposals were unveiled last week to widen access to higher education across the nations of the British Commonwealth by using new and well-established techniques of long-distance learning. The 'University of the Commonwealth for Co-operation in Distance Education' would use correspondence courses, tape recordings and video cassettes, and would seek access to electronic mail and communications satellites to link universities with the aim that "any learner, anywhere in the Commonwealth, shall be able to study any distance-teaching programme available from any *bona fide* college or university in the Commonwealth".

The proposals come from a group of independent education experts set up by the Commonwealth secretary-general last October to report on the potential for Commonwealth cooperation in distance learning. The report, (*Towards a Commonwealth of Learning*, Commonwealth Secretariat, London, £6) says: "Students can, in principle, follow a course from an institution outside their country. Universities can share their teaching through satellites. Audiotapes, video cassettes and discs can be distributed like books." The programme would help colleges to reach new audiences and to widen the range and raise the quality of teaching on campus, and the university would be of particular value for would-be students remote from an educational institution, and for part-time students, the report says.

The proposed university would begin with a small nucleus of about 20 professional staff, building up to about 60 over five years, when it would be able to offer the equivalent of some 20 full degree

courses. The operating budget for the first year is estimated at around £2.4 million, with an annual expenditure of £5-6 million for the first five years, increasing to £8 million thereafter. Teaching material would be built up from the best courses available from open universities and distance-teaching universities throughout the Commonwealth.

The report suggests that where computer technology is required, "major multinational computer companies might well help with the funding of courses and the supply of computer hardware and software on favourable terms".

The subject areas of science, mathematics and technology are singled out as being particularly appropriate for international distance-learning, specifically mathematics and the physical sciences. The report acknowledges that many educational institutions with established distance-teaching techniques have hesitated to introduce science teaching, often because of the difficulty of arranging practical work. It points out, however, that open universities have successfully introduced home science kits and vacation laboratory work. Worldwide demand for applied scientists and technologists, particularly in electronics, information technology and computer studies, made this "one of the most promising areas for development of cooperation". The university would need a location where it could be "assured of government support and would have easy access to Commonwealth resources, with good communications facilities, and with an adequate infrastructure of services".

The proposals will be considered by Commonwealth education ministers later this month and by heads of government in October.

Simon Hadlington

Import of animal viruses opposed after accident at laboratory

Sydney

OPPOSITION to the import of animal viruses into Australia has intensified following an accident at the controversial Australian Animal Health Laboratory (AAHL). A potentially dangerous poultry virus seems likely to have been carried from the laboratory in the form of a human infection.

The accident occurred when a laboratory technician neglected to check that a filter was in place in the apparatus she was using to concentrate a solution of the virus. Live virus sprayed onto her face and hair. After showering twice, she was told she could leave the laboratory but that she might develop conjunctivitis, a symptom of the disease in humans. Three days later the symptoms appeared.

The technician was counselled to keep away from poultry but any possibility that she could have passed on the information arouses strong emotions among Australian farmers. The virus was the Albiston-Gorrie strain of Newcastle disease which devastated Australian chicken farms in

the 1930s. Today the poultry industry employs 25,000 people and is worth A\$640 million a year.

The AAHL facility was designed to provide Australia with the capacity rapidly to diagnose exotic animal diseases. Right from the start it has met opposition because it has not only collected together disease agents from around the country but also imported, for the purposes of research, diseases that do not exist in Australia. Three diseases deadly to pigs have been imported in the past six weeks. Built at a cost of A\$160 million at Geelong, near Melbourne, the facility is thought to be the most advanced microbiological containment facility of its type in the world, and escape of biological agents to be impossible. The New South Wales Farmer's Association will now seek a Federal Court injunction to restrain any further import of live exotic viruses.

According to Dr Bill Snowden, chief of the AAHL, the decision to send the laboratory technician home was made because no case of the virus being transmitted from human to animal, or human to human, has ever been recorded. Snowden regrets that the technician was not given more extensive counselling, as it has come to light that tissues she used to wipe her eyes were thrown into a rubbish bin. The virus could be picked up by wild birds that forage on rubbish tips.

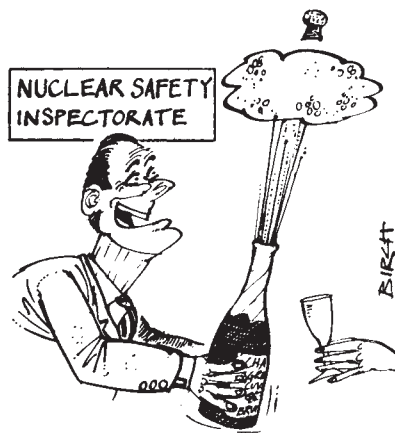
AAHL does not have the facilities or the power to quarantine workers. But Snowden now says, given public reaction to the incident, that he would think twice about letting the person concerned go home if a similar accident occurred. He has ordered an independent inquiry into a number of breaches of AAHL procedure that may have contributed to the accident. Many questions have already been raised: the acting chief of AAHL was not notified until five days after the accident, the chief veterinary officer of Victoria was not notified for a week, and the poultry industry was not notified at all. **Charles Morgan**

Old jobs, new rates

London

BRITISH nuclear industry safety inspectors have been given a pay rise to bring their salaries more into line with what they could get from industry or the private sector, to try to bring a halt to chronic staff shortages.

The independent Advisory Committee on the Safety of Nuclear Installations has



recommended an increase in Nuclear Installations Inspectorate staffing by 50 per cent on present levels, to 150, to cope with present and future workloads. The inspectorate is reviewing the safety of Britain's ageing Magnox reactors, especially any earthquake risk, and it has also been involved in time-consuming reviews of the recently-approved Sizewell B pressurized water reactor.

Kathy Johnston

Major funding for new Canadian centres of excellence

Toronto

IN an effort to reverse the trend of Canadian dependence on imported high technology, the province of Ontario will spend \$200 million over five years to create seven "centres of excellence". Eight Ontario universities and 100 Canadian companies will take part. The competition for money from this new source is intended to stimulate cooperation between Ontario universities and industries.

The seven centres, each a joint venture involving up to six universities and 15 companies, were selected from 28 proposals submitted to Ontario's new "Premier Council", which is directing a \$1,000-million fund supporting science and technology research in Ontario. The proposals were reviewed by a 13-member international panel, advised by 98 external reviewers from Japan, Britain, France, Canada and the United States. John Polanyi, the University of Toronto's recent Nobel prize-winner in chemistry, called the programme "a major and imaginative undertaking".

Polanyi will be a principal investigator at the Centre for Advanced Laser and Lightwave Research at the University of Toronto, which will concentrate on novel lasers and sources, lasers in medicine and related areas. The other planned centres are the Institute for Space and Terrestrial Science at York University, the Centre for Integrated Manufacturing at McMaster University, the Centre for Groundwater Research and the Centre for Information

Technology, both at the University of Waterloo, the Centre for Materials Research at Queen's University, and the Telecommunications Research Institute of Ontario at both Carleton University and the University of Ottawa.

Each centre will be subject to a rigorous review after 30 months and, according to the provincial government, if the mandate is not followed, money will be pulled out. J. Fraser Mustard, chairman of the evaluation panel, says that the success of each programme will be judged by the extent to which the research contributes to industrial growth in Ontario. In announcing the initiative at a press conference on 19 June, Ontario Premier David Peterson said that, "we don't want research sitting on the laboratory table. We want it applied to the real world." Peterson also said that he believed the new centres will allow the creation of a "critical mass" of the best available minds.

R.W. Nicholls, director of York University's Centre for Research in Experimental Space Science, says the new centres will "support long-term fundamental and applied research, but must spin off the research into industry. There will be a visible industrial benefit in a few years time." Nicholls said the project was "big.... As far as York is concerned, we are talking about \$40 million over 5 years. That's several times more than the current budget of our entire faculty of science. This project will certainly put Ontario on the world map."

Russell McNeil

US military loses control of computerized databases

Washington

THE US administration appears to have abandoned its controversial attempt to place control of all computerized databases in the hands of the Department of Defense. With the unanimous passage of the Computer Security Act by the House of Representatives last week, measures to ensure civilian control, worked out with the reluctant cooperation of the administration, are ready to be installed. The act has still to make its way through the Senate but its prospects appear good.

The fight between Congress and the administration over who should control unclassified computer data has now been running for more than two years. President Reagan took up arms in September 1985, when he approved a National Security Decision Directive that gave responsibility for setting security standards for unclassified but "sensitive" information to the National Security Agency.

Opposition grew into outrage when the then National Security Advisor, Vice-Admiral John Poindexter, followed up with a memorandum in October 1986, appearing to extend the area covered by the directive to all unclassified non-government databases. That meant that access to databases carrying journal abstracts, to electronic networks for exchange of scientific data and to any other information service using telecommunications or computers could be controlled by the military. The ultimate aim appeared to have been to find ways to eliminate database and network users who might pass on unclassified scientific information to the Soviet bloc.

Poindexter's memorandum was not fated to survive long. Civil rights groups, library and scientific associations and even bankers, who saw the possibility of military control of electronic cash networks, joined in voicing opposition. In Congress, Representative Jack Brooks (Democrat, Texas) labelled the "expansion of military influence into society" as "Big Brotherism". After Poindexter lost his job because of his involvement in the illegal operation to finance Nicaraguan rebel groups through arms sales to Iran, the memorandum was rescinded. But that did not still congressional opposition. The administration sensed it was on the losing side and compromised by offering its support for a new Computer Security Act.

The urgent need for improved security is recognized by the act's chief sponsors, Representatives Dan Glickman (Democrat, Kansas) and Brooks. But it is not the Soviet Union they are worrying about. The federal government is totally dependent on electronic information systems for

social security, tax and census records and, Glickman says, there is ample evidence of a "disaster waiting to happen" given the potential for "unauthorized access and disclosure, fraudulent manipulation and disruption".

The act gives responsibility for developing security policy on unclassified information to the National Bureau of Standards. It establishes a research programme for developing computer security stan-

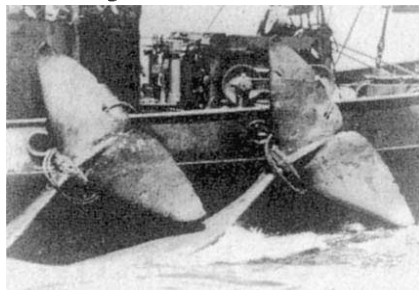
dards and guidelines and requires that training be given by the bureau to computer security officials in each government agency. Private computer databases will not be interfered with. No authority is given to the federal government to control non-government systems and access to government information continues to be guaranteed by the Freedom of Information Act.

Control of classified information remains with the National Security Agency and, in a compromise with the administration, it is given a role in advising the National Bureau of Standards on technical security guidelines. **Alun Anderson**

IWC cracks down on activities of "scientific whaling" nations

London

THE International Whaling Commission (IWC) is cracking down on nations that conduct whaling in the name of science, with a series of votes designed to force Japan, Iceland and Korea into halting their killing of whales.



Whaling ship with catch.

At its annual conference in Bourne-mouth last week, the 41-member IWC was clearly split. "It was very much a case of whaling nations versus nonwhaling nations", says Dr Ray Gambell, IWC secretary. On a 3 to 1 majority, member nations passed various resolutions designed to tighten up the criteria for "scientific" whaling, which is widely seen as providing a convenient smokescreen for a continuation of whaling in the face of IWC's moratorium.

Japan, for example, had announced its plans for a scientific harvest of nearly 900 whales annually in the Antarctic, a programme that clearly provoked nonwhaling members.

The United States put forward a resolution to close the science loophole by strengthening the definition of "scientific" whaling, imposing strict criteria on killing for research. Britain then proposed a resolution criticizing Japan's scientific programme and calling for the Japanese government to put a stop to it. The motion was passed by a substantial majority, as were similar resolutions against scientific whaling by Korea and Iceland.

Norway's proposals for an annual sci-

entific harvest of 200 minke whales were not formally submitted to the IWC and were therefore not the subject of a vote, but the majority does not support the Norwegian plans.

The US resolution also calls for the scientific committee of the commission to review proposed programmes to make sure they would provide useful information about whale stocks that could not be obtained by non-lethal methods.

Even without the support of the scientific committee, the governments concerned would still have a right to issue a permit for scientific whaling, as the IWC cannot legally interfere. But after last week's vote, if a country defies the resolution, the United States could impose sanctions, stopping fishing in its territorial waters and preventing fish imports from the countries concerned. This is seen as a strong deterrent, as the commercial value of these fisheries considerably outweighs the value of the whaling industry.

Whaling nations have argued that their research programmes would be scientifically useful, providing data on whale populations. The programmes would generate income through the sale of whale meat to Japan.

Iceland was particularly upset by the IWC resolutions, arguing that the action is contrary to international law. Early in the meeting, Iceland threatened to quit the commission. Japan's commissioner Tatsue Saito called the outcome "a great disappointment" and said he is concerned about the future of the "conservationist" IWC.

Ironically, if the IWC action brings to a halt all "scientific" whaling, other more legitimate research could also be affected. "Our whale sighting research in Antarctica depends on the presence of Japanese vessels for refuelling", says Gambell. "It would be a serious loss if we had to stop whaling research altogether."

Kathy Johnston

Suicide among women chemists

SIR—For women chemists in the United States, the suicide rate is five times the national average for white women. For men chemists, it is double the national average. These results have been confirmed. They are rising.

Chemistry is the only physical science for which such statistics are available. It is the only one with a high enough proportion of women to allow the comparison between male and female suicide rates. Its homogeneity makes it well-suited to sociological study. For these reasons we are investigating the causes of suicide in this group. Our aim is suicide prevention.

Our method consists of soliciting cases by word-of-mouth. Every precaution is being taken to see that nobody is further hurt. Strict confidentiality is being observed. Nobody's name will be published and nobody will see the raw data except myself and my consultant, a well-known suicidologist in the San Francisco Bay area.

So far I have collected 64 cases. Only 9 of these are women. In order to get a comprehensive picture, I need more cases among women.

If any of your readers knows of any suicides among women chemists or biochemists in the United States, I should be glad to hear from them with as many details as possible.

MOLLY GLEISER

*Suicide Prevention Among Scientists,
1920 Bonita Avenue,
Berkeley,
California 94704,
USA*

Sober view of AIDS

SIR—In using total absolute differences as a criterion for assessing parameter-combinations, Malcom Rees (*Nature* 326, 343–345; 1987) makes an unsuitable choice of statistic as a given discrepancy is not modified in any way by the corresponding expected value. Thus a discrepancy of 10 compared to an expected value of 20 is treated equally with a discrepancy of 10 compared to an expected value of 200. A more reasonable fitting procedure is to use maximum likelihood, and if this is applied to his figures, taking due account of the coarseness of the data, the progressive right-hand truncation and the left-hand truncation at zero (without which a normal can hardly be entertained as a survival function) it suggests a mean of 6 and standard deviation of 2 rather than the values of 15 and 3 proposed by Rees.

The likelihood is, however, extremely flat and very little reliance should be placed on our, or indeed any other, estimates, based on these data. This is hardly

surprising when one considers the small numbers involved and the contradictory evidence provided. The later cohorts provide the greater numbers of AIDS (acquired immune deficiency syndrome) victims but because of progressive censoring provide less and less information for a fit and the final cohort provides no information at all.

Even if the table of probabilities that Rees provides could be taken as providing a reliable model, which it cannot, it would still be extremely difficult to produce reliable estimates based upon it. A relatively small number of AIDS cases, exhibiting at the very least the variability of a Poisson process, are being inflated using tail area probabilities to produce estimates of human immunodeficiency virus (HIV) infection whose variances will be considerably greater than their expected values.

In fact, in view of the paucity of the available data, the possible methodological difficulties of applying results from transfusion-related cases to victims of AIDS in general and the problems of fitting suitable models, it is perhaps wisest to admit that in this matter, estimates are little more than guesses.

JOHN DAGPUNAR
ANDREW LAWSON
STEPHEN SENN

*Department of Mathematics and
Computer Studies,
Dundee College of Technology,
Bell Street,
Dundee DD1 1HG, UK*

Western hypocrisy

SIR—Your comments on the environment and foreign aid (*Nature* 326, 539; 1987) make a telling point about the hypocrisy behind much of the rhetoric about preserving the environments of developing countries. Without something more tangible than exhortation they will be quite unable to respond. But you perpetrate two factual errors and miss a number of important points.

First, the country containing the Amazon rain-forest (Brazil) is not the largest developing country (China is bigger), and it is far from being the poorest in *per capita* terms; according to World Bank tables, it had a gross national product per head of \$1,700 in 1984. There were 77 developing countries below Brazil in the league, down to Ethiopia with only \$110 per head. It is true, of course, that there are millions of Brazilians living in absolute poverty, as the distribution of income is extremely unequal, and by European standards the country is poor.

The real need, for Brazil and other developing countries, is not foreign aid to preserve their environments but a radical overhaul of the terms of trade and finance that keep them poor. Yet it appears that the United States and Britain, both indus-

trial consumers of coffee, recently engineered a collapse in the price that will reduce the export earnings of Brazil. At the same time, the two governments were reported as declining yet again to do anything to lift the impossible burden of debt faced by Brazil. Britain's Chancellor, Nigel Lawson, said it was purely a matter between Brazil and the commercial banks, who still insist on payment at inflated interest rates.

Lawson's other unhelpful suggestion was yet another agreement with the International Monetary Fund (IMF). Brazil's poor need this like a hole in the head; the IMF's cure is for them far worse than the disease, because it inevitably leads to a further drop in their living standard. Brazil's government has now signalled that it can take no more of this.

New thinking is desperately needed on the part of the governments of the richer countries. Their failure to respond to the developing countries' deepening crisis will mean the death not just of the tropical forests but of many millions of human victims of collective greed and neglect.

MICHAEL SPENCER

*34 Bayham Road
Sevenoaks TN13 3XE, UK*

Pre-embryos?

SIR—Your editorial statement (*Nature* 327, 87; 1987) "The fact is that a fertilized human egg is as much deserving of being called an embryo as is a fertilized frog's egg" betrays a lack of understanding of the strategy of early mammalian development. The fertilized frog's egg at the onset of gastrulation has given rise to the embryo and nothing but the embryo, so that embryonic development can indeed be traced back to the 1-cell stage. The fertilized human egg has to manufacture the life-support systems required to support intrauterine existence before it can safely produce an embryo. At the onset of gastrulation, when the human embryo is first formed, it involves less than 1 per cent of the tissue derived from the fertilized egg. The remaining 99 per cent has gone to form the placenta and other nutritive and protective structures. To refer to the previous mammalian stages — be they 1-cell, 8-cell or blastocyst — as embryos is therefore no more (maybe less) appropriate than to refer to them as placentae. As I have tried to indicate previously (*Nature* 320, 570; 1986), appropriate terms for the totality of tissue derived from the fertilized mammalian egg up to the time of gastrulation include conceptus, zygote, ovum, pre-embryo and pro-embryo.

ANNE McLAREN

*MRC Mammalian Development Unit,
Wolfson House,
(University College London),
4 Stephenson Way,
London NW1 2HE, UK*

Solid-state physics resurgent

A conference in California last week may have failed to explain superconductivity of the new kind, but it is none the less important on that account.

WOULD this be another great occasion, with special sessions carrying on into the small hours, like the much-remembered meeting in New York last March of the American Physical Society? One veteran of that occasion told of having lined up outside the closed doors an hour and a half ahead of time, and of having gone home weakly, soon after midnight, long before the special session on ceramic superconductivity ended.

Comparisons with Woodstock (something to do with modern music) are apt; there have been many pop-concerts since that time in the early 1960s, but no repetition of the first occasion. How could there be, when there is only one discovery to celebrate?

Even so, last week's meeting, technically an international workshop of new mechanisms of superconductivity, was a grand occasion. It had begun life with a discreet invitation to fourteen people to gather for an exchange of views and speculations, and had ended up filling the largest hall of the hotel which, while still within the city limits, is the furthest away from the campus that has put this Californian town on the map. More than that, the meeting had the three people who have given their names to the conventional BCS theory of superconductivity — Bardeen (John Bardeen, from the University of Illinois at Champaign-Urbana), Cooper (Larry R. Cooper, now at the US Office of Naval Research) and (briefly) J. R. Schrieffer (something important in industry).

Delegations from elsewhere, but especially from the Soviet Union, were lionized. Although the great V. L. Ginzburg failed to materialize, his colleague Lev Gor'kov from the Lebedev Institute in Moscow did, together with five compatriots. Californians have been much surprised to discover that Russians are charming people (and that they smoke so many cigarettes).

The Japanese were also prominent, willing like the rest to speculate about new mechanisms of superconductivity, but also eager to chance their arms on how it may be possible to make wire out of ceramics. (Extrusion through high-temperature dies seems to be a favourite recipe.) Fair play, others were just as ready to guess when and how there will be some application for the ceramic superconductors, but they were listened to as carefully as were the Japanese.

One striking feature of this cosmopolitan occasion was the prevalence of foreign accents among those masquerading as representatives of US laboratories. The Soviet delegation would have sensed at least three Russian accents in the voices of speakers invited from US laboratories (but one is that of an Israeli citizen). IBM seems strong on immigrants from Britain.

Friendliness was universally conspicuous. That, by itself, is something of a phenomenon. When people say that they find meetings valuable in opening their minds to new ideas, they often mean that they have found them useful as a means of joining a different clique. But on this occasion, people went out of their way to explain that they have fellow-feeling for all their colleagues. "We are a band of brothers", said one, in public. When asked by a journalist where the discovery of ceramic superconductivity stands in relation to that of DNA, another completely missed the point by saying that solid-state physics has nothing like that cut-throat competitiveness to spoil its collective bonhomie.

The truth is that the people at the conference felt in their bones that developments in the past few months have liberated them from dullness. "Condensed-matter physics" is what, after all, it has been called. People have been inured to the pitying tolerance of those in other fields who know that the calculations that can be done are either inexact or uninteresting, and that the important problems in the field are beyond computation.

Of course, the field does encompass a few well-known phenomena of some character. Birefringence (as with calcite) is a little piece of magic, about which too many are too blasé. The superfluidity of liquid helium is another. The notion that a ring of lead or of bismuth, if cooled in liquid helium, will sustain an electrical current once set going indefinitely, say for a year or a century, would have long since seemed magical if the phenomenon had been more familiar. That the same may now be possible merely with the help of a little liquid nitrogen seemed to many at this meeting to have put solid-state physics on the map.

The general opinion, nevertheless, was that there is not yet a theory to account for what is happening in these ternary oxides (see page 14 for a more formal account). Even so, the motto of this meeting may yet turn out to be something along the lines of

"BCS is dead; long live BCS". Put crudely, on the tercentenary of Newton's *Principia*, the immediate objective has been to "save the appearances".

The conventional BCS theory is a largely successful proof that, in a superconducting metal, pairs of electrons near the energetic top of a conduction band will be so affected by their mutual interaction with the vibrations of the ionic lattice in which they move that they will paradoxically attract (not repel) each other, turning from pairs of fermions into bosons which can then occupy the same quantum levels and, with a little luck, form a Bose-Einstein condensate with the properties of a superfluid.

Electrons are then taken, two by two, out of the conduction bands of the material and are allowed, as bosons, to sink to energy levels several electron-volts below. Since there is no Pauli exclusion principle to keep them apart in energy, their energy is merely a reflection of the external temperature. The intuitively missing ingredient is this picture is the function of the quanta of the lattice vibrations, called phonons naturally enough.

The excitons, plasmons, magnons and the like that have occupied last week's theoreticians are means of making BCS come true, but at liquid-nitrogen temperatures. Unfortunately, the attempts to calculate the problem *ab initio*, by starting with the wave functions which are for the time being the best guesses as to how electrons are distributed in the layer structures of a curiously valence-indifferent family of oxides, were either unconvincing or hard to understand.

Meanwhile, K. Alex Müller of IBM's Zurich laboratory, whose work (published last December) first suggested that the ternary perovskite oxides might be liquid-nitrogen superconductors, has no clear account to give on how he came to that development. He tried nickel oxides first, over several years, but they did not work. Copper oxides, he says, were a natural alternative, a remark that points to one of the ingredients missing from the whole of these proceedings, the sense that chemists have of the way in which the electron affinity of different atoms affects the properties of the structures in which they are emplaced.

But it could yet be that the discovery of the ceramic superconductors is the best thing to have happened to solid-state chemistry, not physics. **John Maddox**

Agricultural genetics

Towards insect-resistant plants

Robert Shields

Most plants are resistant to most pests and pathogens: if they were not the landscape would be far bleaker than it is today. Nevertheless, insects and pathogenic microorganisms do cause damage to many agricultural and horticultural plants. Pest control can be achieved by breeding resistant plants or by applications of agrochemicals, but as chemical control is expensive and can be environmentally undesirable, considerable effort is being focused on improving the plants' own defences against attack. On page 33 of this issue¹, Vaeck *et al.* approach this goal using a genetic engineering technique.

The starting point for this work was the observation that the bacterium *Bacillus thuringiensis* makes a polypeptide active against various insects. During sporulation of the bacteria, δ -protoxins are produced and deposited as parasporal inclusions. Once ingested by a susceptible insect, the δ -protoxin is solubilized and cleaved by proteases in the insect gut to form the active toxin. *B. thuringiensis* δ -toxins are considered attractive biocontrol agents as they have no known effects on non-target organisms and have been used as safe biological insecticides for more than 20 years. Different strains of the bacterium make toxins with different specificities, for instance var. *israelensis* is most active against Diptera (such as mosquitoes and blackflies); var. *kurstaki* and var. *berliner* against Lepidoptera (moths); and var. *tenebrionis* and var. *san diego* against Coleoptera (beetles). The relationships between the various toxins is unclear as only a few have been fully characterized and others differ immunologically and in amino-acid sequence^{2,3}. In their new work¹, Vaeck *et al.* cloned subfragments of the gene encoding the var. *berliner* δ -toxin and inserted them into plants using vectors based on the tumour-inducing plasmid of *Agrobacterium tumefaciens*. The regenerated plants express the foreign genes and produce sufficient insecticidal protein to protect them from damage caused by larvae of tobacco hornworms. The newly acquired gene is stable and is transmitted in a mendelian fashion to subsequent generations.

Leaving aside such issues as ecological acceptability and whether plants expressing the *B. thuringiensis* δ -endotoxins taste funny is the question of how useful genetically engineered single-gene resistance will be. As yet no field resistance of insects to commercial formulations of the *B. thuringiensis* toxins have been reported, despite years of use. These preparations are complex mixtures of bacterial

spores, δ -endotoxin and β -exotoxin, and have relatively short half-lives after foliar application. None of these factors favours the emergence of resistant pests. But under intense selection (for instance in stored grain treated with toxins), resistant pests emerge remarkably rapidly⁴, as they also do if conventional insecticides are widely used over extensive areas for many years. Could the widespread use of plants depending on a single gene encoding pest resistance invite similar problems?

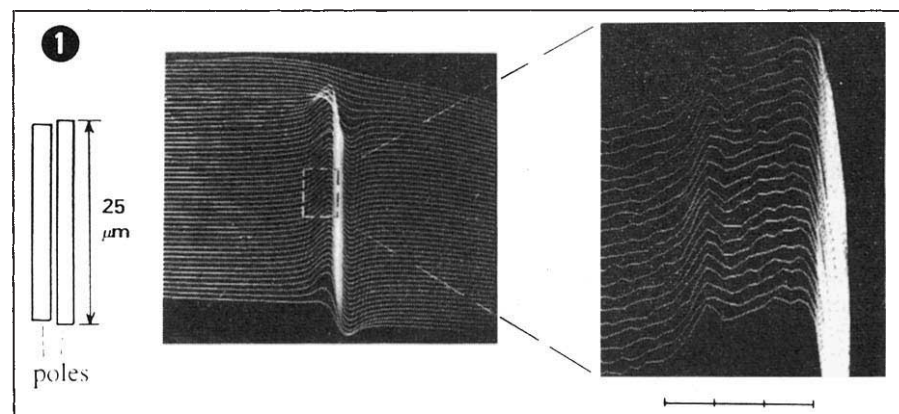
The answer may be to combine several resistance genes in a single plant. As mentioned above, most plants have a degree of resistance against most insects. These

resistances may be architectural (for example, glandular trichomes on the leaf surface, which secrete sticky substances to trap insects or antibiotics to poison them), or they may depend on complex plant secondary metabolites that act as feeding repellants or that inhibit egg laying. Such resistances are likely to be polygenic in character, more appropriate targets for the conventional breeding strategies that have been so successful in the past rather than for genetic engineering. Nevertheless, considerable effort is now being expended in identifying single clonable genes that can be used to protect plants from insect damage.

A promising set of such genes is the wound-induced protease inhibitors. After certain plants are wounded mechanically or by a chewing insect, 'protease inhibitor inducing factor' (PIIF) is released and this induces the systemic synthesis of several protein protease inhibitors. These

As magnetic materials become increasingly important in information technology, witnessed by the development of high-density data-storage devices such as magnetic and magneto-optic digital tapes, so too does the need to understand the physics involved. Y. Martin and H. Wickramasinghe (*Appl. Phys. Lett.* 50, 1455-1457; 1987) at IBM now report a deceptively

the field (the variation in density of field lines). A static field would be hard to detect, but an oscillating field from the sample, or induced in the probe by an alternating current, will cause the probe to vibrate. A laser heterodyne interferometer detects the vibration amplitude ($\sim 2 \text{ \AA}$) with great accuracy — better than $10^{-4} \text{ \AA}$. This can be translated directly into field



simple-looking device that maps the magnetic field of a sample. Their method, with a resolution better than $1,000 \text{ \AA}$, avoids the complications of sample preparation of other techniques.

Figure 1 shows the field pattern of an IBM 3380 tape head. The magnetic field is detected by a magnetized probe, tapered to 0.1 μm , suspended 0.2 μm (or less) above the sample surface (Fig. 2). The force felt by the tip is proportional to the gradient of

gradient. An $x-y-z$ piezoelectric base scans the sample in a raster pattern so that a picture is built up (scale bar, 3 μm).

Higher resolution, for which the tip is suspended $100 \pm 50 \text{ \AA}$ above the sample, requires a slightly different approach. Oscillations of the needle near its resonant frequency (where it is at its most sensitive) are induced by a piezo-crystal at its base. The magnetic (and van der Waals') interactions effectively alter the Hooke's-law spring constant of the needle, changing the force constant. Feedback to the sample base, proportional to the field gradient, adjusts the needle elevation and hence keeps the frequency constant. Effects of the sample's topography can be overcome simply by reversing the magnetic field and subtracting the second image so obtained from the first.

Roland Pease

inhibitors are directed against insect and microbial proteases rather than those of the plant, and they may help to defend the plant against attack by interfering with digestion in the insect gut. The genes encoding some of these protease inhibitors have been cloned and transferred to other plants. In the case where the gene encoding the potato proteinase inhibitor II is transferred to tobacco, the foreign gene is induced when the tobacco plant is wounded or detached leaves are treated with PIIF — despite the absence of any homologous proteinase-inhibitor gene in tobacco^{5,6}. Interestingly, no induction is observed in transgenic plants unless the 3' untranslated region of the potato gene is retained. Some of the sequence in this region is the same as that of the plant extensin gene, a gene induced on attack by microbial pathogens. So it seems likely that genes providing resistance against

insects are part of a larger panoply of related genes encoding plant-defence proteins.

It remains to be seen how effective the *B. thuringiensis* δ -toxin genes or the protease-inhibitor genes will be in improving plant resistance to insects. They may be a panacea, but there is a danger that if too much reliance is placed on too few resistance genes (whether they are normal resident genes or those introduced by genetic engineering) we shall be storing up trouble for the future. □

1. Vaecck, M. *et al.* *Nature* **328**, 33–37 (1987).
2. Aronson, A.I. *et al.* *Microbiol. Rev.* **50**, 1–24 (1986).
3. Herrnstadt, C. *et al.* *Biotechnology* **4**, 305–308 (1986).
4. McGaughey, W.H. *Science* **229**, 193–195 (1985).
5. Sanchez-Serrano, J.J. *et al.* *EMBO J.* **6**, 303–306 (1987).
6. Thornburg, R.W. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 744–748 (1987).

Robert Shields is at the Colworth Laboratory, Unilever Research, Sharnbrook, Bedford MK44 1LQ, UK.

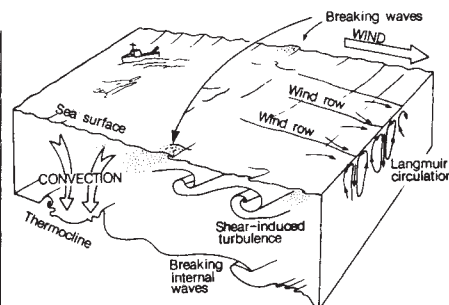


Fig. 1 Summary of the different components of the air-sea fluxes. The net heat flux is the sum of the latent and the sensible heat flux, the infrared or long-wave flux and the solar or short-wave flux. Latent heat flux occurs when the sea surface cools by evaporating; sensible heat flux is the turbulent or convective heat transfer of heat driven by a difference in the temperature of the air and the sea surface. Momentum is transferred by surface stress and by breaking waves. Mass is transferred during evaporation and precipitation and also by processes such as entrainment into the ocean of air bubbles and ejection into the air of droplets.

Oceanography

Mixing in the upper ocean

Robert A. Weller

How do conditions in the atmosphere influence the ocean, and vice-versa? On page 48 of this issue, S. A. Thorpe and A. J. Hall describe how, using a small catamaran towed off to one side and ahead of a research ship, they could observe the distribution of air bubbles and temperature anomalies in an undisturbed region of the upper boundary layer of the ocean (see their Fig. 1 on page 49). Careful analysis shows that, at a given depth, maximum bubble densities coincide with positive temperature anomalies. The authors interpret this finding as evidence for downward transport of warmer, bubble-rich water from near the surface into the region below and suggest three processes that could be responsible for such transport.

Thorpe and Hall's description of their experiment and the ensuing data analysis will be read with great interest by oceanographers and others concerned with the physics of the upper boundary layer of the ocean. More generally, the work represents new and significant progress in understanding how variability in the ocean couples to that in the atmosphere. Although there is a working understanding of how momentum, heat, and mass are exchanged at the air-sea interface (Fig. 1), there has been only limited success in identifying the physical processes that carry those quantities away from the interface into the upper ocean. Solar radiation penetrates the upper tens of metres of the ocean but decays exponentially with depth, losing approximately 80 per cent of its energy in the first 10 metres. Breaking surface waves also inject momentum and

mass just beneath the surface (Melville, W.K. & Rapp, R.J. *Nature* **317**, 514–516; 1985), but air-sea exchanges occur essentially at or within a few metres of the sea surface. Thus, it is mixing processes such as those discussed by Thorpe and Hall in this issue that link the interior of the ocean to the top 10 metres or so, and involve its large heat capacity (4 times that of air) and mass (260 times that of the atmosphere) in atmosphere-ocean interactions.

In the surface boundary layer of the ocean, the vertical profiles of mean horizontal velocity, temperature and

density indicate the strength of the vertical mixing processes. For example, if the transport of heat away from the sea surface was a slow process, then a strong vertical temperature gradient would be found near the surface where heat flows from the atmosphere to the ocean. It is common, however, to find nearly uniform profiles of temperature, velocity and density in the surface boundary layer. This depth independence indicates the presence of efficient mixing processes; and the ocean's surface boundary layer is often called the mixed layer, because of its vertical homogeneity. To determine what processes are responsible for the rapid mixing has been a challenge; but various candidates have been considered (Fig. 2). Under conditions of heat flow from the ocean to the atmosphere (net-negative

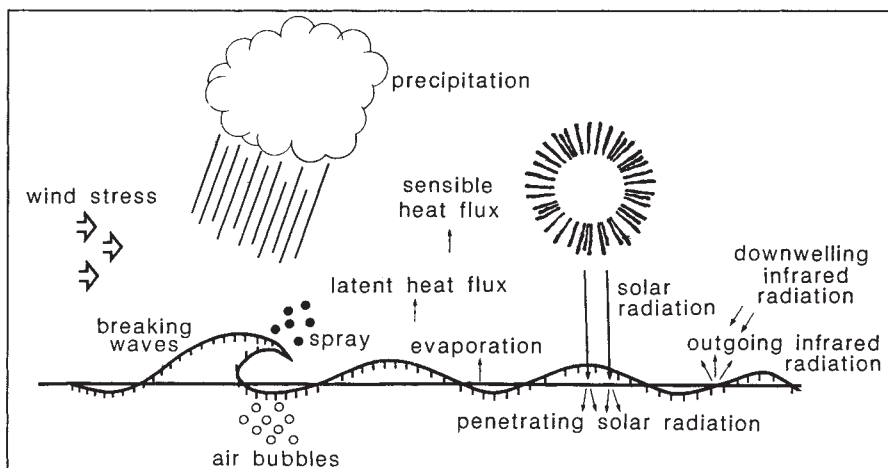


Fig. 2 Some of the mechanisms thought to contribute to vertical mixing in the upper-ocean boundary layer. Cooling the surface results in gravitational instability and during the convection that results the more dense fluid sinks toward the base of the boundary layer. Breaking waves carry surface water and air bubbles below the surface. Large counter-rotating eddies called Langmuir cells are found aligned nearly parallel to the wind; they carry surface fluid together into convergences and then down into the interior. Fluid may also overturn and form eddy-like structures perpendicular to the wind as a result of instability in the horizontal flow (shear or Kelvin-Helmholtz instability). (From Thorpe, S.A. *Nature* **318**, 519–522; 1985.)

heat flux), deepening of the mixed layer has been observed, and convection is recognized as the responsible process (Lacombe, H. *et al.* *Nature* **227**, 1037–1040; 1970). Under net-positive heat flux, the mixing mechanisms work against the tendency of the surface heat flux and penetrating solar radiation to create a stable stratification near the surface. In this situation it is suspected that relatively large scale, coherent flows such as Langmuir cells are important in mixing, but it has been exceptionally difficult to verify such a hypothesis experimentally.

Part of the difficulty in looking for evidence of mixing associated with Langmuir cells, breaking waves or other large eddies is that the measurements must be made close to a moving interface populated with energetic surface waves. Surface waves introduce large-amplitude, high-frequency noise and, in particular, complicate the task of making accurate direct measurements of three-dimensional flow near the surface. Also, because the mixing is rapid, the anomaly signals of water carried down from the surface are small, and are easily confused with signals that result from process other than vertical mixing. For example, Langmuir circulation draws water from the surface where heat exchange occurs, into convergent regions where it is carried down into the mixed layer; the anticipated temperature anomaly associated with the downwelling flow is small, approximately 5 mK.

Thorpe and Hall overcome these difficulties by innovative methods. Their catamaran, which rides the waves, permits measurements near the interface without disturbing the fluid below. They observe the air bubbles that are produced by breaking waves and find that they rise very slowly and thus are reliable passive tracers of currents. If the wind is stronger than about 3 m s^{-1} , breaking surface waves produce discrete bubble clouds; stronger winds produce proportionately more bubble clouds. Under various wind conditions, then, bubbles are created near the surface and are excellent independent indicators of vertical transport. Thorpe and Hall were thus able, using combined bubble and temperature measurements, to develop conclusive evidence for downward transport of surface water. Their techniques can be used to observe Langmuir circulation, injection of fluid by breaking waves and other large-scale structures associated with vertical mixing. It remains now to monitor when such processes are present, to determine the forces that drive such flows, and to quantify the transport associated with them. Once that is done, there is a good chance that dynamically correct models of upper ocean mixing can be constructed. □

Robert A. Weller is at the Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

Experiment ahead of superconductor theory

Berkeley

THEORETICIANS attending last week's international workshop on novel mechanisms of superconductivity (22–26 June) were largely agreed only that the mechanism accounting for superconductivity at liquid-helium temperatures will not suffice for the new ceramic materials. But even that opinion, that the phonon-mediated interaction between electron pairs which is the basis for the standard BCS theory must be supplemented by something else, is not unanimous.

Several groups now agree that substitution of ^{18}O for ^{16}O causes no change of T_c , the onset temperature of superconductivity. This absence of an isotope effect is why most people believe phonon-based (BCS) mechanisms are unlikely. But some theorists argued last week that it is possible that the isotope effect may be suppressed; thus A. Zettl (University of California, Berkeley) showed that substitution of Cu and Ba in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ causes no shift of T_c , but oxygen substitution in a La–Sr superconductor lowers T_c by about 1 K. On the reasonable assumption that similar mechanisms operate in both the Y–Ba and La–Sr compounds, this suggests to some that phonons should not be forgotten altogether.

All remaining theories are based on pairing interactions mediated, in some way, by various electronic motions: the collective motions of electrons (plasmons), local polarization of electron orbitals (excitons) and spin fluctuations travelling through the lattice (magnons). What became apparent as last week's meeting progressed was that, while theorists are still arguing whether various mechanisms can even produce the observed high-temperature transitions, experiments are revealing a variety of surprising and sometimes subtle effects beyond the scope of present theories.

The dependence of T_c on the proportion of oxygen in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ is well known. In a series of experiments at Argonne National Laboratory (I. W. Schuller), samples were quenched from different sintering temperatures, freezing in a controllable oxygen fraction. The transition temperature of the resulting products varied smoothly, until the superconductivity disappeared at $\delta = 0.5$.

Simple counting of valence electrons shows that with less oxygen than this, no Cu^{2+} is present. X-ray diffraction also reveals that this is the point at which the structure changes from orthorhombic to tetragonal. But in a variation on this experiment (B. Batlogg), oxygen was removed at room temperature or below by using a zirconium getter. Here, too, T_c decreases, but in a stepwise fashion, from 90 K to 60 K and then to 30 K. These results are not inconsistent; when oxygen is

removed at low temperature, the lattice cannot easily readjust, and indeed Batlogg reported no change in structure at $\delta = 0.5$. Findings such as these suggest a remarkable sensitivity of the superconductivity to details of the lattice structure; Batlogg believes that the exact placement of oxygen vacancies, not simply their overall number, is important. In a similar vein, Russian work reported by N. Zavariskii (Soviet Academy of Sciences) showed that superconductivity can be destroyed by slight neutron irradiation.

Other experimental results promise to be of more direct use to the theorists, although the 'dirty' nature of the ceramics complicates measurement of their microscopic properties. A fundamental quantity, yet to be pinned down, is the energy gap Δ that represents the binding energy of the superconducting electron pairs. Zettl showed measurements of the absorption edge in infrared reflectivity indicating a gap of $2\Delta/kT_c = 2.4$, lower than the standard BCS ratio of 3.5. But S. Maekawa (Tohoku University) showed that the absorption edge in strongly anisotropic materials (such as these ceramics) is substantially lower than the gap energy; Zettl's result may be consistent with BCS.

The other standard way of measuring the gap, by tunnelling conduction, has been attempted, but because the ceramics have such irregular surfaces, no reliable values have emerged. The values reported were nevertheless reasonably close to the BCS ratio, perhaps telling against resonant valence bond (P.W. Anderson, Princeton) or superexchange (J. Appel, University of California at San Diego) mechanisms, both of which suppose a small or vanishing gap.

This meeting was unusual for the almost complete absence of speculation about superconductivity at temperatures above 100 K. Most comments on the various reports were sceptical. Workers at Westinghouse (A. Braginski) had partially substituted fluorine for oxygen in the Y–Ba superconductor and, although they found anomalous behaviour of the resistivity at about 220 K, they found no indication of a complete loss of resistance.

There was also controversy about a report that the compound $\text{Y}_3\text{Ba}_5\text{Cu}_4\text{O}_{14}$ is superconducting at 90 K (M. Beasley, Stanford). Beasley presented X-ray diffraction spectra purporting to show that his sample consisted of a single phase, called 3-3-6. Others, however, argued that the sample had more than one phase, including that known as 1-2-3 which is known to be a superconductor. The significance, according to Beasley, is that the 3-3-6 phase does not incorporate the Cu–O chains said by some to be central to the superconducting mechanism.

David Lindley

Mineralogy

Evolution of cooling magmas

Richard Wilson

THE Skaergaard intrusion in East Greenland has long been the textbook example of closed-system fractional crystallization of basalt, in which magma forced into the continental crust deposits minerals of varying composition during cooling (Fig. 1). At Skaergaard, the thermally convecting magma was thought to have become increasingly iron-rich and silica-depleted as crystallization progressed. In a recent paper, R.H. Hunter and R.S.J. Sparks (*Contr. Miner. Petrol.* **95**, 451–461; 1987) disagree with this view, arguing that on the contrary, after magnetite started to precipitate in the Lower Zone, the liquid became depleted in iron and enriched in silica. This challenges one of the classic features of Skaergaard and, as iron has a strong influence on magma density, it has important consequences for magma dynamics during crystallization.

The problem concerning the evolution of the liquid composition arises because no rocks representing the liquid composition at any stage are preserved. L.R. Wager and G.M. Brown (*Layered Igneous Rocks* 1–588; Oliver & Boyd, Edinburgh, 1968) chose a chilled-margin sample as the parental magma composition, and then used mass-balance calculations involving estimates of the volumes of the different zones in the Layered Series, assuming closed-system fractionation with the final liquid represented by the Sandwich Horizon rock. Their calculations (Fig. 2) imply that the liquid showed extreme iron enrichment but no silica enrichment until about 98% of the magma had crystallized. McBirney proposed even greater enrichment of iron in the melt (*Nature* **253**, 691–694; 1975) on the basis of melting experiments that determined the compositions of liquids trapped between accumulated minerals.

Hunter and Sparks now point out the problems of these techniques for estimating contemporary liquid compositions, and emphasize that no volcanic rocks

comparable in composition to the calculated Skaergaard liquids have been found. They argue that the magma must have reached its maximum iron enrichment when magnetite started to precipitate and that this magma contained 50–52% SiO_2 ; also that precipitation of magnetite gabbro containing about 44% SiO_2 caused the silica content of the melt to increase. After magnetite entry, the liquid differentiated from ferrobassalt, giving basaltic andesite, then icelandite and then rhyolite, a common sequence in similar, tholeiitic volcanic provinces. The liquid line of descent therefore enriched the iron content only during the earliest stages of crystallization, after which the trend was one of silica enrichment and iron depletion. If the Hunter and Sparks liquid compositions are correct, a larger volume of residual rhyolitic differentiates must have been produced than are preserved in the intrusion. One possibility is that these evolved rocks have been removed by erosion; another is that Skaergaard may have fed the rhyolitic magma to a volcano.

If the Skaergaard magma became depleted in iron and enriched in silica after the start of magnetite precipitation, its density would have decreased in the later stages of fractionation. Although this means that plagioclase crystals would have been able to sink from about the Middle Zone, it would still have been buoyant in the earlier stages of fractionation. This buoyancy therefore remains an important objection to the argument that the layering in Skaergaard developed by crystal settling. But these density considerations help to explain the presence in the Middle Zone of large blocks very rich in plagioclase derived from the Upper Border group, which previously were believed to have been buoyant in the Middle Zone parental magma.

Another important consequence of decreasing density with fractionation after magnetite precipitation starts is that

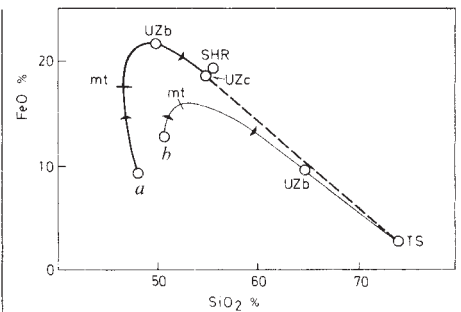


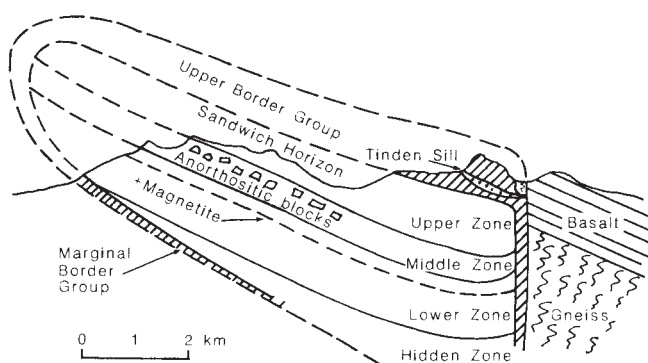
Fig. 2 Plots of Fe as FeO against SiO_2 for the calculated liquid compositions during the evolution of the Skaergaard intrusion. Solid line shows the trend calculated by Wager and Brown from the chilled margin EG4507 (a) to Upper Zone C (UZc). Dashed line shows trend observed in filter pressed liquids. Fine line shows trend calculated by Hunter and Sparks for Lower Zone B liquid with 52% SiO_2 , from the chilled margin KT 39 (b, from Hoover, J. D. Yb. Carnegie Instn Wash. 77, 739–743; 1978). mt, Entry of magnetite at top of Lower Zone B; UZb, Upper Zone B; SHR, Sandwich Horizon Rock, a cumulative from a rhyolitic liquid; TS, final liquid composition from Tinden Sill.

residual magma released during crystallization would be buoyant and able to rise to the top of the magma chamber. This evolved, buoyant liquid could have become trapped in the Upper Border Group rocks and may explain why they are more evolved than their Layered Series equivalents. The rise of evolved liquid would also enable the Skaergaard magma to have become compositionally zoned, and Hunter and Sparks suggest that by the time the Middle Zone was crystallizing, the magma in contact with the roof was strongly zoned from rhyolite to basaltic andesite, overlying a larger volume of convecting ferrobassaltic magma.

Many of the questions posed by Hunter and Sparks concerning the evolution of the Skaergaard magma could well apply to other fractionated plutonic rocks. They emphasize that evolved liquids with high iron-to-magnesium ratios do not need to contain exceptionally high absolute iron to produce iron-rich cumulate rocks.

The Skaergaard Intrusion has long been considered the classic example of closed-system fractional crystallization by crystal settling, showing extreme iron enrichment in both the rocks and the liquid. A.R. McBirney and R.N. Noyes (*J. Petrol.* **20**, 487–554; 1979) had already argued persuasively that crystal settling did not play a major role in the evolution of Skaergaard. Now it seems that the liquid did not become exceptionally enriched in iron, that the magma may have become compositionally zoned, and that the chamber may have been open during part of its history. The Skaergaard intrusion is still classic, but it is also becoming an example of a discarded hypothesis. □

Fig. 1 Simplified schematic cross section through the Skaergaard Intrusion. Broken lines above the horizon represent the assumed original extent of the intrusion. Magnetite is found above the broken line labelled + magnetite. Based on Fig. 16 of Irvine, T.N. in *Origins of Igneous Layering* (ed. Parsons, I.) 185–245 (Reidel, Dordrecht, 1987).



Richard Wilson is at the Geologisk Institut, Universitetsparken, 8000 Aarhus C, Denmark.

Topology

The hyperbolic phoenix

Ian Stewart

ONE of the most fascinating things that can happen in mathematics is the discovery that two parts of the subject, previously thought to be unrelated, have surprising new connections. When one of those parts has largely been written off as a dead end, the other is of central importance to present-day research, and the combination promises a breakthrough at a fundamental level, fascination gives way to high drama.

Mathematical dramas, however, are often somewhat low-key. The drama that I have in mind really began in the late 1970s, when William Thurston of Princeton began to circulate a set of lecture notes in which he applied non-euclidean geometry to three-dimensional topology. My excuse for writing this article now, rather than then, is that Thurston is only now beginning to publish his ideas in the official literature: the first instalment of a seven-part series has just appeared (*Ann. Math.* **124**, 203–246; 1986). But that is just an excuse: my true reason for writing is that such a seminal development cannot be allowed to go unremarked. The climax of the drama, in any case, may be yet to unfold.

The story goes back to the middle of the previous century, when mathematicians were just starting to realize that there were many more things in mathematics than had been dreamt of in their philosophy. There were new geometries, new arithmetics and new algebras; creative freedom was in the air. The task of mathematics was still to reflect the workings of nature, but people were beginning to understand that they should not take that message too literally.

One major triumph of nineteenth-century mathematics was the invention of non-euclidean geometry, in which Euclid's axiom of parallels is disobeyed. Another was the complete classification of the possible topological types of surfaces. The two themes were linked by the discovery that every type of surface possesses its own natural geometry. The topological theme waxed with the incisive work of Henri Poincaré, and modern algebraic topology was born. The geometrical theme, in comparison, appeared to peter out. But appearances can be deceptive.

Interest in the topology of surfaces was originally aroused by Bernhard Riemann in 1851. His immediate objective was to clarify the nature of multivalued complex functions, but the ideas soon gained wider importance. Riemann constructed his surfaces by taking a series of concentric spheres, making slits in them and joining

them up in crosswise fashion. In 1877 William K. Clifford published a simpler, but topologically equivalent, description: a Riemann surface is a sphere with holes bored through it. Felix Klein suggested an alternative model, a sphere with handles stuck on. By 1874 Klein had proved that not just every Riemann surface, but every closed two-sided surface, can be so represented. Klein also noted that there exist one-sided surfaces, such as the projective plane and the Klein bottle.

Surfaces are also crucial to non-euclidean geometry. A vivid way to realize a system of geometry other than



Euclid's is to consider the geometry of a curved surface. Euclidean concepts, such as straight lines, are replaced by analogues, here geodesics, or shortest paths. On the surface of a sphere, 'straight lines' are great circles, and any two such meet. That is, there are no parallels. Geometers distinguished three main types of geometry, which they called elliptic, euclidean and hyperbolic. Elliptic geometry is essentially that on a sphere, or positively curved surface. Euclidean geometry is that of a plane, which has curvature zero, whereas hyperbolic geometry is that on a negatively curved surface, shaped like a saddle near any point. One example is the pseudosphere, an object resembling a trumpet whose mouthpiece is at infinity. A better way to visualize hyperbolic geometry is to consider the interior of a circle, and represent a 'straight line' by an arc of a circle that cuts the boundary at right angles. This is the disk model of the hyperbolic plane.

That was the state of play at the turn of the century. Every surface has a unique natural geometrical structure, and all surfaces can be classified. But the ideas, methods and proofs were all specific to

two dimensions. It seemed unlikely that anything similar would work for spaces of dimension three or higher. Higher-dimensional spaces, or manifolds, however, were becoming ever more important in mathematics.

By the 1960s topologists were beginning to understand manifolds of dimension five or higher well enough to answer the questions of immediate concern. By the late 1970s the work of Andrew Casson and Michael Freedman had paved the way to a similar understanding of four-dimensional manifolds. But in three dimensions there were more questions than answers. At this point Thurston revived the idea of natural geometrical structures, which had previously seemed hopeless in three dimensions or more, pointing out that they might provide a powerful new weapon. But the technical difficulties involved were considerable.

Three-dimensional space has its own versions of elliptic, euclidean and hyperbolic geometries — but these are no longer exhaustive. Thurston showed that five other types of geometry must also be considered. He also realized that it was asking too much to expect an entire three-dimensional manifold to model itself on just one such geometry. Instead, the manifold must be cut into several pieces, each piece with its own type of geometry, but differing from one piece to the next. This is Thurston's geometrization conjecture: every three-dimensional manifold has a natural decomposition into geometrical pieces.

This is a very surprising conjecture. A topological manifold is a floppy thing, free to be kneaded and distorted at will, like a lump of pizza dough. Geometry is much more rigid, like a cooked pizza crust. Any link between the two must reflect very deep features of the structure of manifolds. The conjecture remains unproved, but evidence in its favour is accumulating, and it is now known to be true for a very wide range of manifolds, in particular a class called Haken manifolds. Thurston's series of papers will describe what is currently known. In particular, for seven out of the eight geometries, the problem is essentially solved. The most difficult, challenging and interesting case is three-dimensional hyperbolic geometry: Thurston's main theorem proves that every 'atoroidal' Haken manifold carries a hyperbolic geometrical structure.

The geometrization programme met with an early success, when it contributed to the settling of a long-standing problem called the Smith conjecture. In the 1940s Paul Smith proved that if a three-dimensional sphere is mapped to itself by a periodic topological transformation then the set of fixed points is topologically a circle. (A transformation is periodic if, when repeated several times, it sends each point back to where it started. A fixed

point is one that is transformed to itself.) Smith asked: can the circle of fixed points be knotted? He conjectured that it cannot. The proof was completed in 1978, as the result of the combined efforts of at least a dozen mathematicians working in several different branches of mathematics (*The Smith Conjecture* eds Morgan J. W. & Bass, H. Academic, London, 1984). Strictly speaking, the conjecture is false for certain 'wild' transformations, but it is true if, for example, the transformations are sufficiently smooth. Other applications of Thurston's ideas, for example to complex dynamics, attest to the fertility of the methods.

The mathematics used to attack the geometrization conjecture is varied and original. It includes geometry, group theory, algebraic topology, complex function theory and foliations. Thurston remarks that "There has been a rather long gap since I first formulated the geometrization conjecture and announced and discussed its proof for Haken manifolds. . . . Part of the difficulty of writing about it has been that there is a lot of background material that feeds into the proof at least in the informal sense, if not in the logical sense. This background material is nice in itself — it seems to be an intrinsic part of the geometric theory of three-manifolds — but it is daunting because of its quantity". Many of the difficulties stem from the need to break the manifold into pieces: this is often required

for technical reasons during the proofs, even in cases where the final result is that the entire manifold carries a single geometrical structure. In consequence the behaviour of the geometrical structure near boundaries must be kept under control. "The hard part of the proof", says Thurston, "is the geometric analysis of all possible hyperbolic structures on a manifold with boundary in order to find limits to the way the geometry is distorted as the hyperbolic structure is varied".

The power of Thurston's ideas may be gauged from the fact that, if the geometrization conjecture is true, then the celebrated Poincaré conjecture, that the only simply connected three-dimensional manifold is the three-sphere, follows immediately. (For other recent developments on this problem and further information see my recent News and Views articles in *Nature* 320, 217–218; 1986 and 325, 579–580; 1987.) But Thurston's ideas go far beyond this, promising, if substantiated, to open up the whole structure of three-dimensional manifolds, one of the last great frontiers of topology. Meanwhile, interest in three-dimensional hyperbolic geometry in its own right has also revived. As the needs of mainstream research change, supposedly 'dead' subjects can rise like the phoenix from their own ashes. □

Ian Stewart is at the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

DNA chemistry

How the double helix breathes

Maxim Frank-Kamenetskii

It has long been known that the storage of genetic information in the form of a DNA duplex permits its safe transfer from generation to generation. One reason is that the most reactive sites of the DNA bases are buried inside the double helix, making them inaccessible from the outside. This decreases DNA vulnerability in terms of damage by inherent (OH^- and H^+ ions) and foreign (countless mutagens and carcinogens) agents. But to what extent does this structure actually protect DNA? New results that are reported on page 89 of this issue¹ by Guéron, Kockoyan and Leroy support an unorthodox answer to this question.

Controversy about this issue has never ceased since the discovery of the double helix. Most people believe, on the basis of the hydrogen-exchange data, that there is a rather high probability (about 10^{-3}) of each base pair opening (breathing) because of thermal motion, so that the protective effect of the DNA structure is not very significant. This orthodox view has found its way into the textbooks². A

minority (including myself) has insisted, however, that the protective effect is quite significant and that the probability of base-pair opening is about 10^{-5} .

The latter view is based on theoretical estimations and studies of DNA modification by formaldehyde³. Because the hydrogen-exchange data have always seemed more direct, these arguments could not rival the orthodox view. Now the situation has been changed by the results of Guéron *et al.*, who show in this issue¹ that the traditional interpretation of the hydrogen-exchange data is incorrect. Their revised estimation leads to the very figures that have long been regarded as unacceptable.

Guéron *et al.* used a traditional strategy of probing DNA breathing by the hydrogen-exchange method^{2,4}. Their technique is based on the natural assumption that the imino protons (N3 for thymine and N1 for guanine), which are buried the deepest inside the double helix, can be exchanged only from open base pairs. The rate-limiting stage of the overall

exchange process for the DNA duplex may be either the base-pair opening or the exchange chemistry. These two possibilities can be distinguished by adding a catalyst. If it affects the overall process, the rate-limiting stage is chemical, whereas if the exchange is independent of the catalyst, then the base-pair opening is the rate-limiting stage. In the latter case the rate of the overall exchange process determines the lifetime of a base pair. In the former case the probability of base-pair opening can be estimated by dividing the exchange rate for the duplex by the exchange rate for free nucleotides.

The orthodox interpretation of the hydrogen-exchange data rests on the fact that the exchange of imino protons is catalyst-independent^{2,4}, an assertion based on traditional isotope methods, which cannot resolve very fast exchange rates. Guéron *et al.* studied short duplexes by magnetic resonance, which allows the exchange rates of individual imino protons to be followed over a wide timescale. They found a threshold concentration of the added catalyst above which the exchange rate of the imino protons becomes strongly catalyst-dependent, and they conclude that exchange chemistry rather than base-pair opening is the rate-limiting stage of the imino-proton exchange. This result calls for a complete revision of the conventional interpretation of the hydrogen-exchange data. Guéron *et al.* show that the base-pair lifetime is in the 10-millisecond range, whereas the opening probability is about 10^{-5} , in full agreement with my earlier estimation⁵ using the formaldehyde method.

The data of Guéron *et al.* resolve another long-standing question: does the double helix bend smoothly or is it a series of straight segments connected by hinges? If the double helix were interrupted by open base pairs on average every 100 base pairs, the hydrodynamic and other macromolecular properties of DNA could be explained without invoking the inherent bending flexibility of the double helix⁶. The data of Guéron *et al.* strongly support the view^{3,6} that base-pair breathing does not contribute to the macromolecular properties of DNA, which can be accounted for by smooth bending of the double helix.

Guéron *et al.* show that low catalyst concentrations do not affect the exchange rate of the imino protons in DNA duplexes because these duplexes contain an inherent catalyst. But what is the catalyst? The answer is completely unexpected: it is the complementary base. In the open state, mutually complementary bases are retained near each other, and those without imino groups serve as catalysts for their partners (adenine for thymine and cytosine for guanine).

Overlooking this self-catalysis of the duplex had resulted in an overestimation

of the base-pair lifetime and mimicked a large base-pair opening probability^{2,4}. This had led to the most puzzling feature of the conventional view, an unreasonably large lifetime (up to fractions of a second) of the open state^{2,4}, which even caused speculation about solitons (see my News and Views article to be published next week) travelling along DNA.

The revised estimate of base-pair lifetime and opening probability give a perfectly reasonable figure of 10^{-7} s for the open base-pair lifetime. Thus, after years of uncertainty, the method of hydrogen exchange has finally produced a complete set of reliable DNA double-helix

breathing parameters: base-pair lifetime 10^{-2} s; open base-pair lifetime 10^{-7} s; and base-pair opening probability 10^{-5} . □

1. Guéron, M., Kochoyan, M. & Leroy, J.-L. *Nature* **328**, 89–92 (1987).
2. Cantor, C.R. & Schimmel, P.R. *Biophysical Chemistry* Ch. 22 (Freeman, San Francisco, 1980).
3. Frank-Kamenetskii, M.D. in *Structure and Motion: Membranes, Nucleic Acids and Proteins* (eds Clementi, E., Corongì, G., Sarma, M. & Sarma, R.) 417–432 (Adenine, New York, 1985).
4. Mandal, C., Kallenbach, N.R. & Englander, S.W. *J. molec. Biol.* **135**, 391–411 (1979).
5. Manning, G.S. *Biopolymers* **22**, 689–729 (1983).
6. Wilcoxon, J. & Schurr, J.M. *Biopolymers* **22**, 2273–2321 (1983).

Maxim Frank-Kamenetskii is at the Institute of Molecular Genetics, Academy of Sciences of USSR, Moscow 123182, USSR.

Origin of life

Back to the RNA world — and beyond

Geoffrey North

In his 1968 paper *The Origin of the Genetic Code*¹ Francis Crick, speculating on how the complex biochemical mechanisms of protein synthesis could have originated, wrote "It is tempting to wonder if the primitive ribosome could have been made entirely of RNA", adding later "Possibly the first 'enzyme' was an RNA molecule with replicase properties". These words, recalled by Alan Weiner in his summary of this year's Cold Spring Harbor Symposium on Quantitative Biology*, seemed almost uncannily prophetic at the end of a meeting where attention focused on the recent discovery that RNA molecules can have true enzymic activities. The list of such RNA enzymes (or ribozymes) is growing, and with it the diversity of reactions they can catalyse. At the same time, speculation is coming back into fashion among respectable molecular biologists, and plausible schemes are being constructed for some of the key developments in the origin of life, which give a more concrete form to the ideas that flowed from the pioneers of the 1960s and which, most importantly, can at least in principle be assessed experimentally.

The attraction of RNA for those seeking a rational account of the origin of life is as a possible way out of the classical 'chicken-and-egg' problem posed by the existence in all known forms of life of distinct sets of informational and functional molecules, locked in a seemingly interminable cycle of interdependence. How could such a cycle have arisen? A molecule that was both informational and functional might be the answer and RNA, already known in the 1960s to have non-informational roles in protein synthesis, seemed the most likely candidate for such

a molecule. As Crick remarked in 1966², "tRNA looks like Nature's attempt to make RNA do the job of a protein". The possibility that RNA might have had a catalytic role in a pre-protein world was strengthened by the recognition that many coenzymes are nucleotides (coenzyme A, for example) or use bases possibly derived from nucleotides (such as thiamin pyrophosphate), leading to speculation in the 1970s that coenzymes might be fossils of the earliest, RNA-based enzymes³.

It should therefore come as no surprise that the discoveries of RNA molecules with catalytic activities comparable to those of protein enzymes has aroused great interest. The pioneer in the field is Tom Cech (University of Colorado), in whose laboratory it was first shown that the 'intervening sequence' (IVS) of the protozoan *Tetrahymena thermophila* ribosomal RNA can spontaneously excise itself from its surrounding 'exon' sequences in the absence of protein. The subsequent work, culminating in the discovery that the excised IVS has catalytic, and thus true enzymic, activities⁴ was discussed in a News and Views article⁵ last year by Frank Westheimer and will not be recounted in detail here.

One of the reactions the *Tetrahymena* IVS was shown able to catalyse, however, is of particular interest in the context of this article: this is RNA polymerization, an obvious prerequisite of an RNA replicase, and is illustrated in the figure. Zaugg and Cech showed that by a series of transesterifications the *Tetrahymena* IVS can convert pentacytidylic acid (C₅) into polycytidylic acids of a range of sizes, up to C₃₀. Although the reactions are essentially disproportionations and do not change the average oligonucleotide length, they closely resemble those

catalysed by RNA polymerases in that in both cases phosphoester bonds are conserved. The specificity for polycytidylic acids lies in the 'internal guide' sequence of the IVS, which is thought to base-pair to the substrate as shown in the figure. Cech has conjectured that this specificity could be altered simply by changing the internal guide sequence, and his optimism is supported by the recent demonstration that the specificity of the endoribonuclease activity of the *Tetrahymena* IVS can be altered in a predictable manner, generating what are effectively a series of RNA restriction endonucleases⁶.

The next step in establishing the feasibility of Crick's speculation that an RNA replicase could be made solely of RNA will be to show that the IVS internal guide sequence can be replaced by an external template, which is replicated during the reactions. Cech has suggested this in an explicit model⁷ and his group is trying to make the model a reality.

If a molecule that can catalyse its own replication provides a way out of the 'chicken-and-egg' problem, Walter Gilbert has emphasized in a News and Views article⁸ the importance of making the distinction between information storage and function early in evolution, because of the divergent structural demands of these two roles. This theme has been taken up by Weiner and Maizels (Yale University), who have brought together several intriguing observations in devising a remarkably convincing scheme for how the transition to protein synthesis might have occurred⁹.

Weiner and Maizels believe that clues to the origin of life are to be found among the viruses, many of which they consider likely to be 'living fossils' of early stages of evolution. In particular they believe that RNA viruses are likely to be representatives of the 'RNA world', and they have noticed the tendency of such viruses to have transfer (t) RNA-like structures at the 3' ends of the genomes. This is sometimes revealed directly in the 3' sequences of the virus genome, but more often by the ability of the viral genomes to react with proteins known to recognize tRNAs. Some plant RNA viruses, for example, are amino-acylated at their ends by specific amino-acyl tRNA synthetases (T.C. Hall, Texas A&M University), and the enzyme that replicates the RNA genome of bacteriophage Q has as essential subunits the tRNA-binding proteins EF Tu and Ts.

Weiner and Maizels suggest that in the RNA world 3' tRNA-like sequences could have 'tagged' genomic RNAs for replication, the tag being recognized by replicases and removed in the conversion of genomic to catalytic molecules. The non-encoded CCA sequence that is added to the 3' termini of tRNAs and some RNA viruses is suggested to have acted as a primitive

*52nd Cold Spring Harbor Symposium on Quantitative Biology, *The evolution of catalytic function*, 27 May–3 June 1987.

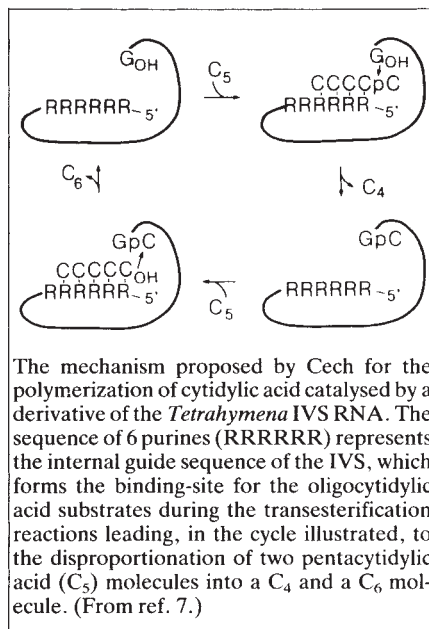
telomere, protecting information at the 3' end of the genome. Weiner and Maizels point out that sequences closely related to CCA occur at the end of modern telomeres. Perhaps most telling of all, the enzyme that in prokaryotes releases tRNAs from their precursors, RNase P, has a crucial RNA component that works as a ribozyme in the absence of protein (S. Altman, Yale University; N. Pace, Indiana University; see also ref. 5), and there is now evidence that an enzyme involved in adding the complement of the CCA-related sequences to telomeres also has an essential RNA component (reported by Greider and Blackburn at the recent Cold Spring Harbor RNA processing meeting).

The next step in the scheme of Weiner and Maizels is the suggestion that the 3' tRNA-like tags could have played a key role in the evolution of protein synthesis if a variant replicase gained the ability to bind amino-acids and 'charge' them by transfer to the 3' tRNA-like end of a genomic RNA molecule, or even a free tRNA-like tag liberated by RNase P. This might have occurred by a series of reactions closely related to those catalysed by the *Tetrahymena* IVS and almost identical to that catalysed by present-day amino-acyl tRNA synthetases. The carboxylate group of the bound amino-acid could displace a mononucleotide from a charged RNA replicase (formed much as in the figure as an intermediate in RNA polymerization) and the amino-acyl NMP intermediate so formed could then react with the 3' end of a tRNA-like tag, thus forming the equivalent of the amino-acyl tRNA substrates of modern protein synthesis. Spontaneous polymerization of charged amino-acids could then generate the first polypeptides, and an RNA that catalysed the polymerization by virtue of having two 'tRNA' binding sites might have been the first primitive ribosome.

The beauty of this scheme is that in the wake of the work of Cech and others, each step seems plausible. Although it is doubtful that the truth of such a scheme could ever be proved, the way of laboratory reconstruction is open to establish it as a real possibility. The versatility of RNA as a catalyst thus becomes a question of great interest. Two new examples of RNA enzymes were reported at the meeting by Olke Uhlenbeck (University of Colorado), both involving the catalysis of reactions chemically distinct from those of the *Tetrahymena* IVS and RNase P. One model derives from the work of Robert Symons (University of Adelaide), who has shown that oligomers of several plant virusoid RNAs undergo self-cleavage to monomeric units, apparently by a common mechanism determined by conserved sequence elements that permit the RNAs to fold into what Symons calls a "hammerhead" secondary structure¹⁰. These same conserved sequence elements

turn up in a small RNA of unknown function found in the newt (satellite 2 RNA), which also undergoes self-cleavage *in vitro* at the site predicted by Symons' model (L. Epstein, Florida State University)¹¹.

Uhlenbeck and colleagues have now synthesized two short RNAs based on Symons' conserved sequence elements and shown that one, a 19-mer, will catalyse the site-specific cleavage of the second (a 24-mer). Turnover shows this is true catalysis and that Uhlenbeck's group has made the smallest ribozyme known to date. The second novel ribozyme described by Uhlenbeck derives from the



The mechanism proposed by Cech for the polymerization of cytidylic acid catalysed by a derivative of the *Tetrahymena* IVS RNA. The sequence of 6 purines (RRRRRR) represents the internal guide sequence of the IVS, which forms the binding-site for the oligocytidylic acid substrates during the transesterification reactions leading, in the cycle illustrated, to the disproportionation of two pentacytidylic acid (C₅) molecules into a C₄ and a C₆ molecule. (From ref. 7.)

observation made in the laboratory of Aaron Klug (Laboratory of Molecular Biology, Cambridge) that lead ions bind at a specific site to yeast tRNA^{Phe} molecules and at the appropriate pH will catalyse cleavage of the tRNA backbone¹². Uhlenbeck reported that tRNA^{Phe} can be split into two halves, one of which will bind a lead ion and catalyse the cleavage of the other half. This ribozyme is of particular interest for two reasons: one is that it is a metalloenzyme, and metals might have been important in extending the catalytic powers of primitive ribozymes; the second is that the structure of yeast tRNA^{Phe} with a bound lead ion has been determined¹², which is particularly valuable as RNA is notoriously difficult to crystallize.

Assuming that self-replicating RNAs could have existed, and plausible schemes envisaged for their evolution to the modern systems of genetic information storage and expression, would the problems of the origin of life be solved? That the answer to this question is emphatically no was made clear by several talks at the meeting. In particular, Leslie Orgel (Salk Institute), Alan Schwartz (Nijmegen), Gerald Joyce (Salk Institute) and Stanley Miller (University of California, San Diego) argue

that RNA must be an evolutionarily advanced molecule¹³. Their argument is based on the seemingly insurmountable barriers in the way of the *de novo* development of self-replicating RNA in the prebiotic soup.

One problem is posed by the complex mixtures of sugars predicted to exist in the prebiotic soup. Perhaps the most severe difficulty is that of enantiomeric cross-inhibition. The termination of chain elongation during replication that results from the incorporation of nucleotides with D-ribose rather than the L-ribose sugar of all modern RNA¹⁴. As a way out of this dilemma it is suggested that RNA was preceded as an informational molecule by a polymer with a different kind of sugar-phosphate backbone. Joyce and colleagues expect that many of the problems faced by RNA could be avoided if the substrates for oligomerization of the first informational molecules were glycerol-derived acyclonucleoside diphosphates. These are prochiral, and although they gain a chiral centre on oligomerization, molecular-dynamics simulations by Michael Levitt (Weizmann Institute) predict a flexible (pyrophosphate) backbone permitting incorporation of monomers in either syn-L-like or anti-D-like conformations without chain termination. As a bonus the duplex of such polymers is predicted to become progressively less stable with increasing length, so that replication of information based on base complementarity can occur but the informational molecules do not become trapped as stable duplexes.

If the meeting was haunted by the pre-science of Crick, it was dominated by the ethos of Theodosius Dobzhansky, whose statement that "nothing in biology makes sense except in the light of evolution" was quoted to effect on several occasions. Perhaps the strongest message of the meeting, made most forcibly in the speculations of Weiner and Maizels, is that the key to turning on that light might be any one of the myriad specific details which molecular biologists are so rapidly accumulating, but which in themselves may appear to have no great general significance. □

1. Crick, F.H.C. *J. molec. Biol.* **38**, 367 (1968).
2. Crick, F.H.C. *Cold Spring Harb. Symp. quant. Biol.* **31**, 3 (1966).
3. White, H.B. III *J. molec. Evol.* **7**, 101 (1976).
4. Zaug, A.J. & Cech, T.R. *Science* **231**, 470 (1986).
5. Westheimer, F.H. *Nature* **319**, 534 (1986).
6. Zaug, A.J., Been, M.D. & Cech, T.R. *Nature* **324**, 429 (1986).
7. Cech, T.R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4360 (1986).
8. Gilbert, W. *Nature* **319**, 618 (1986).
9. Weiner, A.M. & Maizels, N. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
10. Forster, A.C. & Symons, R.H. *Cell* **49**, 211 (1987).
11. Epstein, L.M. & Gall, J.G. *Cell* **48**, 535 (1987).
12. Brown, R.S., Hingerty, B.E., Dewan, J.C. & Klug, A. *Nature* **303**, 543 (1983).
13. Joyce, G.F., Schwartz, A.W., Miller, S.L. & Orgel, L.E. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
14. Joyce, G.F. *et al. Nature* **310**, 602 (1984).

Geoffrey North is Deputy Biology Editor of Nature.

Neurobiology

Arachidonic-acid metabolites as second messengers

Stuart Bevan and John N. Wood

INTRACELLULAR second messengers (cyclic nucleotides, calcium, diacylglycerol and inositol polyphosphates) are known to mediate a wide variety of regulatory extracellular signals involved in functions as diverse as cellular growth control and neurotransmission. The report by Piomelli *et al.* on page 38 of this issue¹ provides strong evidence that eicosanoids (arachidonic-acid metabolites) can be added to the list of second messengers. In this article, three groups at Columbia University and the University of Texas show in an elegant series of experiments that an eicosanoid, probably 12-hydroperoxyeicosatetraenoic acid (12-HPETE), acts as a second messenger at a synapse in the marine mollusc *Aplysia californica*. Specifically, this compound is responsible for the inhibitory effects of FMRFamide on 5-HT (serotonin)-induced depolarization of *Aplysia* sensory neurons. (See the discussion² in News and Views by Jennifer Altman earlier this year, which contains a schematic diagram of second-messenger regulated neural pathways.)

Arachidonic acid, when released from esterified stores in the cell membrane, gives rise to various biologically active oxidation products, collectively called eicosanoids. Two main enzymatic pathways have been defined: the cyclooxygenase cascade catalyses the production of prostanooids; and the lipoxygenase enzymes synthesize leukotrienes, hydroxyeicosatetraenoic acids (HETEs and HPETEs) and lipoxins³. Although the modulatory effects of eicosanoids on cardiovascular, reproductive and immune functions have been the focus of much interest over the past two decades, there has been little information available on any neuronal effects of eicosanoids except for the action of prostanooids on peripheral sensory neurons.

The clues that led Piomelli *et al.*¹ to investigate eicosanoids as potential second messengers originated in their demonstration that histamine application to cerebral ganglia of *Aplysia* evoked arachidonic acid release and breakdown⁴. Histaminergic neurons exert inhibitory effects on abdominal ganglia, and stimulation of these inhibitory neurons also results in arachidonic acid release with a similar spectrum of oxidation products. These observations prompted an analysis of the role of eicosanoids on a better-defined system. 5-HT depolarizes *Aplysia* sensory neurons by indirectly closing a class of

potassium channels by a mechanism that involves cyclic AMP-dependent phosphorylation⁵. The report in this issue shows that the neuroactive peptide FMRFamide (Phe-Met-Arg-Phe-NH₂) antagonizes the effect of 5-HT by increasing the probability of the potassium channels remaining open, although this action does not involve the cyclic AMP phosphorylation cascade. Arachidonic acid itself mimics the action of FMRFamide, and inhibitors of lipoxygenase, but not of cyclooxygenase activity, block the effect of FMRFamide. With the discovery of the lipoxygenase products 5- and 12-HETE both in synaptosomes and in identified cell bodies, the unstable hydroperoxy-precursors of these compounds were tested for their ability to mimic FMRFamide. 12-HPETE shows such an activity and the authors conclude¹ that this, or a related compound, mediates the action of FMRFamide. Therefore, two second-messenger systems converge to act in opposite ways on a single type of membrane channel.

It may be particularly significant that eicosanoids can pass freely across cell membranes so that they may have a role as both intracellular and extracellular signals. This property also suggests a novel means of trans-synaptic modulation of neuronal activity, which Piomelli *et al.* outline in their article. At some synapses, nerve impulses can lead to a prolonged subsequent increase in the postsynaptic response to a presynaptic input, termed long-term potentiation (LTP)^{6,7}. Both pre- and postsynaptic mechanisms have been implicated in LTP, and some of the best evidence shows that a presynaptic modification leading to increased transmitter release can occur during LTP. But at some synapses, the activity of the postsynaptic cell is also important.

Some of the best-studied examples of LTP occur in the hippocampus. Here, pharmacological evidence suggests that activation of postsynaptic NMDA-type glutamate receptors is important for the development of LTP and that polarization of the postsynaptic membrane can influence the establishment of LTP. Experimental evidence, however, shows that there is a presynaptic component of LTP operating at these synapses. Does this result indicate there are two separate loci for LTP or does the postsynaptic activity somehow feed back onto the presynaptic terminal and, if so, how?

Activation of NMDA receptors is

known to lead to an increase in calcium influx, and an increase in postsynaptic intracellular calcium levels is necessary for LTP. Although mechanisms such as calcium-mediated phosphorylation of postsynaptic components could explain any postsynaptic locus of LTP, it has not been clear how such changes could modify the properties of the presynaptic terminal. Eicosanoids provide a possible link. The elevated intracellular calcium concentration could promote eicosanoid production through activation of phospholipases; eicosanoids could then diffuse through the membrane to act on or in the presynaptic terminal. Such an action could add to purely presynaptic events that may also involve calcium; perhaps eicosanoids are involved here too.

The demonstration of eicosanoid-mediated second-messenger activity in neurons poses the question of what other cell types use similar mechanisms of signal transduction. Recent experiments demonstrating G-protein control of arachidonic acid release in various cell types are consistent with the notion that receptor-mediated eicosanoid synthesis is a general signal-transduction mechanism subserving a variety of functions. Evidence for eicosanoid stimulation of protein kinase C⁸, guanyl cyclase⁹ and calcium mobilization¹⁰ from reticular stores suggests that eicosanoids interact with, or mimic, the effects of other second-messenger systems. In support of this idea, a forthcoming article¹¹ by Schaad, Schoderet and Magistretti demonstrates a role for prostaglandins in mediating the synergism between vasoactive intestinal peptide and noradrenaline in raising cyclic AMP concentrations in the mouse cerebral cortex.

With the use of cyclooxygenase and lipoxygenase inhibitors, the range of systems that use eicosanoids as second messengers should soon be catalogued. It will be considerably more difficult to define the precise site of action of these lipophilic and frequently unstable molecules, and to distinguish truly intracellular actions from effects occurring at the cell membrane. □

1. Piomelli, D. *et al.* *Nature* **328**, 38–43 (1987).
2. Altman, J. *Nature* **325**, 111 (1987).
3. Needleman, P., Turk, J., Jakschik, B.A., Morrison, A.R. & Lefkowitz, J.B. *A. Rev. Biochem.* **5**, 69–103 (1986).
4. Piomelli, D., Feinmark, S.J., Shapiro, E. & Schwartz, J.M. *Soc. Neurosci. Abstr.* **12**, 1340 (1986).
5. Castellucci, U.F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **77**, 7412–7496 (1980).
6. Teyler, T.J. & DiScenna, P. A. *Rev. Neurosci.* **10**, 131–161 (1987).
7. Smith, S.J. *Trends Neurosci.* **10**, 142–144 (1987).
8. McPhail, L.C., Clayton, C.C. & Snyderman, R. *Trans. Assoc. Am. Physicians* **97**, 222–231 (1984).
9. Gerzer, R., Brash, A.R. & Hardman, J.G. *Biochim. biophys. Acta* **886**, 383–389 (1986).
10. Wolf, B.A., Turk, J., Sherman, W.R. & McDaniel, M.L. *J. biol. Chem.* **261**, 3501–3511 (1986).
11. Schaad, N., Schoderet, M. & Magistretti, P.J. *Nature* (in the press).

Stuart Bevan and John N. Wood are at the Sandoz Institute for Medical Research, 5 Gower Place, London WC1E 6BN, UK.

AIDS vaccine predictions

SIR—Coates *et al.* recently suggested that a gag p24 vaccine might prevent the onset of AIDS (acquired immune deficiency syndrome) in human immunodeficiency virus (HIV) carriers¹. Their proposal was based on the well-known observation that clinical progression to AIDS is often associated with a reduction in antibodies to p24, the major HIV capsid protein. Whereas this association has been described in AIDS patients from the United States or Europe²⁻⁶, we briefly report that an absence of antibodies to p24 is not found in African AIDS patients.

Sera positive for antibody to HIV were obtained from 117 Africans (40 AIDS patients, 40 AIDS-related-complex (ARC) patients and 37 symptomless subjects) from Burundi, Central Africa. For comparison, sera positive for antibody to HIV were obtained from 80 French citizens (21 AIDS patients, 13 ARC patients and 36 symptomless subjects). Specific antibody to HIV-core antigens (HIV-core Ab) were detected with a competitive enzyme immunoassay in which a recombinant-DNA-produced HIV core protein is used as antigen (Envacore, Abbott). The entire major core protein p24 is present in this recombinant antigen.

Ninety per cent of AIDS patients (36/40) and ARC patients (36/40) and 97% (36/37) of symptomless seropositive individuals were positive for antibody to HIV-core antigen in the African group. But in the French group, HIV-core Ab was detected in only 24% (5/21) of AIDS patients, 74% (17/23) of ARC patients and 86% (31/36) of symptomless seropositive individuals. The prevalences of HIV-core Ab observed in African AIDS patients and in French AIDS patients were highly different ($p < 10^{-6}$). Similar observations have already been obtained using a radioimmunoprecipitation technique (P. Kanki *et al.* *International Symposium of African AIDS*, Brussels, November 1985).

The clinical signs and the incidence of AIDS in Central Africa are very similar to those observed in the United States and Europe^{8,9}, but the fact that African AIDS patients do not have a significant decrease or loss of antibody to HIV-core proteins is contrary to the hypothesis that the immune response to p24 plays a role in protection against the development of AIDS.

FRANCIS BARIN

Laboratoire de Virologie,
CHRU Bretonneau,
37000 Tours, France

1. Coates, A.R.M. *et al.* *Nature* **326**, 549 (1987).
2. Kitchen, L. *et al.* *Nature* **312**, 367 (1984).
3. Barin, F. *et al.* *Science* **228**, 1094 (1985).
4. Schupbach, J. *et al.* *N. Engl. J. Med.* **312**, 265 (1985).
5. Lange, J.M.A. *et al.* *Br. med. J.* **293**, 1459 (1986).
6. Weber, J.N. *et al.* *Lancet* **i**, 119 (1987).
7. Allain, J.P. *et al.* *Lancet* **ii**, 1233 (1986).
8. Quinn, T.C. *et al.* *Science* **234**, 955 (1986).
9. Mann, J.C. *et al.* *J. Am. med. Ass.* **255**, 3255 (1986).

Protein binding to DNA

SIR—In a recent discussion of protein-DNA interaction, Richmond¹ suggested that a synthetic homologue of the normal *EcoRI* restriction endonuclease recognition sequence could test the hypothesis that the sequence itself might determine the unwinding that allows the specific interaction with the protein. He proposed a sequence in which 2,6-diaminopurines replaced the central adenines in the recognition sequence GAATTC. This analogue would have the normal array of hydrogen bonding donors and acceptors in the major groove and added amino groups in the minor groove.

We have, in fact, synthesized such an analogue as an octadeoxyribonucleotide and tested it as a substrate with the *EcoRI* endonuclease² and the modification methylase³. In addition, we synthesized an analogue with 2-aminopurines at these same positions; this has the added amino groups in the minor grooves but lacks the 6-amino groups in the major groove that have been shown to be hydrogen bond donors to glutamic acid side-chains of the *EcoRI* endonuclease in the enzyme-oligodeoxyribonucleotide structure determined by McClarin *et al.*⁴.

Our experiments indicate that the primary effects on the steady state kinetics of the hydrolysis reaction are due to the introduction of the amino groups into the minor groove, where there is no protein in the crystal structure⁴, and not their removal from the major groove. The rela-

tive specificity constant (k_{cat}/K_M) for the analogue with the 2,6-diaminopurine substitutions is 0.11, whereas it is 0.09 for the 2-aminopurine substitutions. These results support the contention that the addition of an amino group to the minor groove of the central A-T base pairs of the recognition sequence interferes with the activity of the enzyme by preventing the DNA from attaining a conformation required for binding or catalysis. The addition of a methyl group to the 3-position of this same adenine in an alkylation-interference experiment prevented binding of the endonuclease to the recognition sequence⁵. The presence of the 2-amino or 3-methyl groups in the minor groove may prevent the unwinding of the DNA helix to form the Type I neokink⁴ that is necessary to open the major groove to allow the protein access to the otherwise cryptic recognition determinants. Sequence-dependent conformational constraints in the DNA may play an important role in determining the specificity of interaction between proteins and DNA.

RICHARD I. GUMPORT

Department of Biochemistry and
College of Medicine,
University of Illinois,
Urbana, Illinois 61801, USA

1. Richmond, T.J. *Nature* **326**, 18 (1987).
2. Brennan, C.A., Van Cleve, M.D. & Gumport, R.I. *J. biol. Chem.* **261**, 7270 (1986).
3. Brennan, C.A., Van Cleve, M.D. & Gumport, R.I. *J. biol. Chem.* **261**, 7279 (1986).
4. McClarin, J.A. *et al.* *Science* **234**, 1526 (1986).
5. Lu, A.-L., Jack, W.E. & Modrich, P. *J. biol. Chem.* **256**, 13200 (1981).

Reversible evolution?

SIR—Harvey and Partridge (*Nature* **326**, 128; 1987) state that most "... biologists view adaptive evolutionary change as reversible." Nowhere in the report is an attempt made to clarify the ambiguity inherent in their concept of reversibility. It would be unfortunate if the non-biologist readers of this journal now hold the misconception that adaptive evolution is reversible in a literal sense.

Reversibility denotes, *sensu stricto*, the capacity to regress to an ancestral state, duplicating all intermediate states, seriatim. Although Dollo's law (which embraces a concept of irreversibility) is interpreted variously, it remains the common view among biologists and historians alike that, even if intermediate steps are ignored, independently evolved phenomena are rarely indistinguishable. Although point mutations are reversible, as evolving systems become more polygenic (that is, as trends involve ever greater numbers of genes), the probability of returning to a given ancestral state becomes vanishingly small, because this quantity is the product of the probabilities of reversion of each involved gene. In the

context of morphological or behavioural change over hundreds of thousands of generations, the occasional isolated reversion of such microscopic events are lost in the manifold trends of organismal evolution. Or they at least find themselves cast against a novel genetic background.

In an effort to provide a basis for reversibility, the authors offer the familiar example of industrial melanism in the peppered moth, *Biston betularia*. The frequency of melanic forms first increased as pollution from the industrial revolution took hold and then decreased as smoke control measures were enforced, as a function of their visibility to predators. It has been argued that this vignette of adaptive evolution in action is oversimplified. Observed allelic frequencies do not conform to a simple model of predation selection, and differential fitness unrelated to crypsis is likely a complicating factor (Jones, J.S. *Nature* **300**, 109; 1982).

Harvey and Partridge then cite an interesting case of behavioural irreversibility in a hymenopteran genus, and adumbrate that such examples are relatively rare. Evidence of evolutionary irreversibility is, *a fortiori*, ubiquitous. A

few familiar examples: mammals returning to a marine habitat did so without re-evolving gills; hominids coming back down from the trees did not re-evolve non-primate characters; and birds returning to the ground have not been endowed again with dinosaurian characters. Indeed, what is rare is a concise example of reversible evolution.

PATRICK L. WILLIAMS

Department of Chemistry,
Northern Illinois University,
DeKalb, Illinois 60115, USA

Nuclear risk evaluation

SIR—Risk analyses by designers of large risky projects, such as nuclear reactors, calculate the probabilities of accidents with the help of a decision tree, and the likelihood of the failures of each component or safety system is assessed¹. The results are the inputs for a chain of calculations leading to probabilistic risk assessment. The unreliability of the method is well known. Accidents of nuclear power plants that were considered by their designers to have a risk of occurring once in a million years have happened twice in a decade, or less than 10,000 factory-operating years². A comparison of the observed failure frequency of civil engineering is several orders of magnitude greater than the safety to be expected from calculations³. Therefore, the technical risk assessment used to calculate nuclear safety must be replaced by an assessment which uses data from operating experiences⁴.

Previous efforts to use observational data to assess nuclear risks made use of only two data points, the Three Mile Island and Chernobyl accidents, assuming that the accidents are Poisson processes⁴. The reliability of a statistical method depends upon the availability of sufficient observational data, and because there have been few accidents at nuclear power plants, the usefulness of a statistical method has been questioned³. The impasse in the statistical approach results from the fact that the data on smaller accidents or incidents have been dismissed; minor incidents have been considered irrelevant because they did not do much damage.

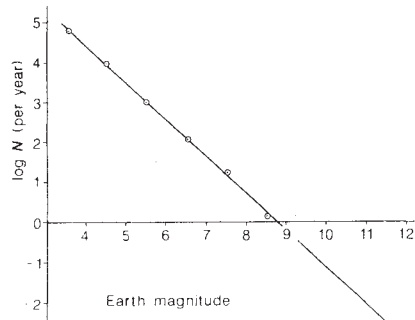


Fig. 1 Earthquake frequency-magnitude plot⁸.

Studying natural catastrophes in history, such as landslides, submarine avalanches, earthquakes, desiccation of a marine basin (up to the size of the Mediterranean Sea), and of meteorite impact, I observed a linear relationship between the magnitude and frequency of the events when these parameters are plotted on log-log graph paper⁵. This relationship has recently been confirmed by statistical studies between crater sizes and impact frequencies⁶. The inverse relationship between magnitude (log energy) and log frequency of earthquakes is well known⁷. I have plotted data supplied by the Swiss Earthquake Service⁸. Millions of earthquakes have occurred since registering began early this century. These earthquakes could be classified on the basis of their magnitude. The data show a perfect linear correlation on a log-log plot (Fig. 1).

Had the statistical analysis of earthquakes been based upon only the few large earthquakes that caused catastrophic damage, the quantitative relationship between magnitude and frequency could hardly have been established. The key to the prediction of reactor catastrophes may thus be to use the data of smaller accidents—what have been called operational disturbances and shutdowns, which are frequently reported by the news media.

The mathematical relationship between the frequency F and the magnitude M caused by incidents should have the general form

$$F = f(1/M)$$

and is

$$\log F = -k \log M + C$$

if it is an inverse log/log relationship like that for earthquakes, where k is the slope of the line, and C is the intercept.

The parameter F is easy to define. Incidents and accidents can be classified into different classes on the basis of their magnitude. The frequency of occurrence of each class can be computed on a factory-year or factory-hour basis.

Several parameters could be chosen to express the magnitude of the accident. One is to use the radioactivity released. But whereas large accidents release a significant amount of radioactivity, smaller incidents may release none, so to use radioactivity release would exclude many small incidents from statistical analyses. Also the amount of radioactive release at Chernobyl is near the upper end of the spectrum; one cannot expect a substantial log/log linear extrapolation of the frequency-magnitude relation far beyond that of Chernobyl.

One useful parameter for measuring natural or man-made catastrophes is a monetary unit. I suggest that we use Swiss francs, a most stable currency, as the parameter M in the frequency-magnitude plot. The usual methods of relating finan-

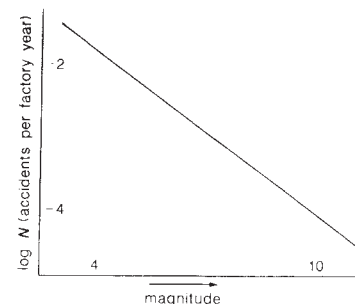


Fig. 2 Nuclear safety plot, suggested relation. Frequency in log factory years, and magnitude in log Swiss francs.

cial losses to incidents and accidents can be used to calculate M . For the smallest incidents, the financial losses may involve only the monetary equivalent of the electricity not produced while the plant is shut down. For small incidents involving minimal radioactivity, additional damage because of the prospect of future cancer deaths should be considered. For a big accident, like Three Mile Island, the very large expenses of dealing with or disposing of a damaged reactor have to be added. For very big accidents, like Chernobyl, there are additional costs, including the 'cost' of actual deaths (to be estimated on the basis of insurance premiums or of damage payment in law suits), damages to property and to agricultural products, genetic damage to unborn children and the values of land and properties that become unusable because of the contamination by radioactivity. The last of these is very important: the monetary value of Swiss properties, for example, is such that an accident releasing the same amount of radioactivity to Swiss towns and villages as that released around Chernobyl would be at least 10 and perhaps even 1,000 times more expensive than it was. We are far from having reached the upper end of the spectrum if the environmental cost is measured in financial terms.

My preliminary impression, judged on the basis of the frequency of shutdowns and the two large accidents, is that the relationship may take the form shown by Fig. 2. A more exact relationship will depend upon the evaluation of data now being requested from nuclear regulating agencies.

KENNETH HSÜ

Geologisches Institut,
ETH-Zentrum,
CH-8092, Zurich,
Switzerland

1. Reactor Safety Study (Nuclear Regulatory Comm. Washington DC, 1975).
2. Wilson, R. *Nature* **323**, 29–30 (1986).
3. Islam, S. & Lindgren, K. *Nature* **322**, 691 (1986).
4. Tiedemann, H. *Nature* **324**, 404 (1986).
5. Hsü, K. J. *Sedimentology* **30**, 3 (1983).
6. Shoemaker, E. M. *A. Rev. Earth planet Sci.* **11**, 461 (1983).
7. Gutenberg, B. & Richter, C. F. *Seismicity of the Earth and Associated Phenomena* (Princeton Univ. Press, 1949).
8. *Erdbeben, Entstehung, Risika und Hilfe* (Schweizerische Geophysikalische Kommission, Zurich, 1986).

In the worst event . . .

Edward Teller

The Medical Implications of Nuclear War. Edited by F. Solomon and R. Q. Marston. National Academy Press, Washington, DC/Wiley, Chichester, UK: 1986. Pp.619. Hbk \$52.25, £44.20; pbk \$40.25, £32.20.

A World Beyond Healing: The Prologue and Aftermath of Nuclear War. By N. Wade. W. W. Norton: 1987. Pp.190. \$15.95. To be published in Britain on 3 September by Sidgwick & Jackson, £12.95.

Lasciate ogni speranza voi ch'entrate — Abandon all hope, you who enter here — is the inscription on the gates of hell in Dante's *Divina Commedia*. As a description of the horrors of nuclear war, those words are apt. *The Medical Implications of Nuclear War*, which contains the papers presented at a conference of the same name, seems to consider them the only proper response to that threat. The author of the foreword to the book establishes the prevailing point of view:

I raise my voice, yell sometimes, what the hell are newspapers for? You're supposed to provide information, real news, and this is the news of the world, the news of the end of the world, print more of it for God's sake before it's too late, I say. Put the nuclear winter up there on the front page every day....

Nevertheless, the conference papers include several careful and objective reports, such as a review of calculations of large urban fires and an assessment of radioactive fallout. Some sections seem less relevant, for instance Part IV of the book entitled "Images and Risks of Nuclear War: Psychosocial Perspectives".

A World Beyond Healing, initially intended as a review of the conference, differs from it. Wade explains his title by saying: "To repair the havoc of a nuclear exchange lasting an afternoon might take the work of generations. Building a new, postnuclear world would not necessarily be impossible. But the old world would be beyond healing". Wade emphasizes a central issue: the rate at which nuclear conflict could develop is unprecedented. He also omits some symposium papers and introduces further pertinent information, including the idea of defence, in discussing other papers.

His stated purpose is to be "impartial", and he succeeds in part. For example, he describes the safeguards against accidental weapon firing, which are hardly mentioned in the related conference reports. However, both books explain that the safeguards change during wartime and that the transition is dangerous. Neither book mentions that the transitional danger is also a deterrent against a first strike.

The most dreadful results of a nuclear war are urban fires and the consequent loss of millions of lives. At Hiroshima,

damage from fire and damage from blast were comparable. As the explosive power of a weapon approaches a megaton, the damage from fire becomes predominant. A conference report presents the physics of large urban fires in a manner as realistic

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Speaking for himself—a victim of Hiroshima.

as can be achieved from known facts and detailed calculations. Three papers estimate the number of casualties, another describes triage in a historical context and a further paper emphasizes the obvious inadequacy of existing medical facilities to provide a peacetime level of care in a post-war setting. No medical procedures or preparations that would moderate the extent of suffering and loss of life are detailed in either book.

The consequences of radioactive fallout are more widespread than those of fire. One participant constructs arguments that under the conditions of nuclear war radi-

ation would have a more lethal effect than is usually assumed. A careful discussion of world-wide fallout reports that the worst case (if nuclear reactors were deliberately attacked) would raise the natural incidence of cancer by no more than one or two per cent. In considering the genetic consequences of radiation, both the conference paper and Wade's book point out that despite detailed and extensive studies statistically significant genetic effects have not been observed among the 19,000 children born to irradiated survivors of Hiroshima and Nagasaki. According to the conference report, current risk assessments would predict "about 240 cases of genetic ill health among the 19,000 children, but none were found".

Seven of the conference reports deal with aspects of the 'nuclear winter', a topic that also receives attention in Wade's book. According to estimates, the amount of smoke that would be generated by the fires in cities is comparable with that now released globally in a year, mostly from forest fires. Under normal conditions, rain removes smoke, but fire plumes and the Sun's heating of dense smoke clouds could well lift the smoke above the moist atmosphere; thus the smoke might escape being rained out for a protracted period. Current calculations predict sizeable temperature drops only if the war is fought in spring or summer. Uncertainties in the amount of smoke emitted and even greater uncertainties in the amount escaping prompt rainout are obvious.

In the past three years, the emphasis in nuclear winter research has shifted from predictions of decreased temperature to concern over lesser changes such as frost or lack of precipitation that might result in failures of harvest. The conference summary states: "The degradation of agriculture by temperature change and the destruction of the world's bread basket — the United States and Europe — will cause tens of millions (possibly even billions) of deaths from starvation". One paper reports that US stored food would feed the nation for about a year and that the United States supplies more than 50 per cent of the grain sold in the world market. Neither book pursues the effect of increased US food storage on post-war hunger or post-war recovery.

Compared with the other dangers, the possible depletion of the ozone layer properly receives less attention. Depletion occurs if the fireball rises into the stratosphere, an unlikely event with current weapons which have smaller yields. A new speculation, based on large amounts of smoke being produced by urban fires and reaching the stratosphere, was presented at the conference; the potential for such smoke to deplete ozone is detailed in both books.

Wade calls the electromagnetic pulse

(EMP) effect "the havoc factor". The effect, generated by a nuclear explosion, damages communication, power supply and electronic devices. The West is more vulnerable to EMP than the Soviet Union. However the Soviets have studied it more extensively. Wade points out that American lack of EMP defence is destabilizing, but both books focus on how, in wartime, damages would reinforce each other. The complementary fact — that civil defence, storage of food and medical supplies, and improvement of dispersed medical facilities also would reinforce each other — is not mentioned.

One symposium paper expresses surprise at "an unusual willingness of participants in the symposium to accept the offered predictive results and speculations at face value", and suggests that "we as scientists will ultimately serve society best by examining every aspect of this issue more critically".

Indeed the horrors of nuclear war are sufficient to make peace of utmost importance. Ignoring defence or any other means to deter war is an error. To insist upon a single and final answer to a complex political problem seems particularly tragic.

In July 1953, the *Bulletin of Atomic Scientists* published an article by J. Robert Oppenheimer which includes this paragraph:

[Defensive] measures ... will mean, first of all, some delay in the imminence of the threat. They will mean a disincentive, a defensive deterrent, to the Soviet Union. They will mean that the time when the Soviet Union can be confident of destroying the productive power of America will be somewhat further off — very

much further off than if we did nothing. They will mean, even to our allies, who are much more exposed and probably cannot be well defended, that the continued existence of a real and strong America will be a solid certainty which should discourage the outbreak of war.

Deterrence has emphasized retaliation in an unbalanced manner. Defence can never provide complete protection, but it could save lives, moderate suffering and, most important of all, it could contribute to deterrence. It is regrettable that the possibility of defence was so pre-emptorily dismissed by the conference.

Both books consider the effects of nuclear war in absolute terms. Neither the practice of science nor the practice of medicine can justify the neglect of partial solutions. If we succeed in stabilizing peace, it will be almost certainly through a series of partial successes in which active defence, passive defence and the preparedness to give help in all possible ways will play important parts in the process.

The Medical Implications of Nuclear War raises the spectre of nuclear war as the final apocalypse. Wade's book does not. The reader who has never considered the terrible effects of nuclear war can find some useful information in *World Beyond Healing*. It is much more difficult to recommend *Medical Implications*. It is not a useful guide for those who seek ways to prevent war or for those who want to contribute to healing and human well-being should such a catastrophe occur. □

Edward Teller is a Consultant at the Lawrence Livermore National Laboratory and a Senior Research Fellow at the Hoover Institution, Stanford, California 94305, USA.

Out of Africa

Derek Ager

Neogene Paleontology and Geology of Sahabi. Edited by Noel T. Boaz *et al.* Alan R. Liss, New York/Wiley, Chichester, UK: 1987. Pp. 401. \$175, £159.65.

MY FIRST reaction to this book was "where is Sahabi?". Not minding making a fool of myself is one of the advantages I have noticed of advancing years, and I asked six colleagues. Fortunately they didn't know either, though I had the least excuse as I worked not far from there. It turns out that Sahabi is in eastern Libya, and we might have been told so in the title.

Question two was "Was it worth producing such an expensive book about the palaeontology and geology of the area?". The answer to that is definitely "yes". Sahabi has a remarkable record of early Pliocene vertebrates, and a great deal of expertise has here been brought to bear on it by a team of American, Belgian, German, French and Libyan specialists. The book is a report of the first four years of an investigation of one site south of Benghazi, where Italian soldiers first found bones in the late 1920s, and is a detailed and well-illustrated account of an almost unique fauna.

The preservation of vertebrate fossils usually depends on the habitats of the animals concerned. Thus by far the most abundant vertebrate fossils are of fish, crocodiles, turtles and hippopotami. This is because these animals are simply asking to be fossilized in the mud of which they are so fond. There is no way that horses and elephants, lions and snakes will be preserved except in very special sites such as fissures and caves. At Sahabi, however, a great variety of land, freshwater and marine forms are found in the channels at the mouth of a great river system.

The fauna has some Eurasian elements and, surprisingly, some connections with that of Central and South America, but generally it is African in character. The link between Europe and North Africa was probably cut by the salty desert of Late Miocene times, and the Sahabi fauna is largely a relict descended from what was there before. Remarkably, almost everything is present: fish, snakes, crocodiles, turtles, monkeys, whales, carnivores, "elephants", "sea-cows", "horses", "giraffes" and "cattle" (using these terms mostly in a very wide sense). There are also some forms which are rarely preserved, such as water-birds and, perhaps most interestingly, hominids.

Also included in the book is an account of the flora and microfossils of the area, together with a summary of the sedimentology and regional geology, and con-



"Ejecting an intruder" — the illustration comes from Alfred Russel Wallace's *The Malay Archipelago*, an account of his travels in the region and of its biology. The book was first published in 1869, and has recently been re-issued by Oxford University Press with an introduction by John Bastin. Price is £27.50, \$35.

clusions about the taphonomy and palaeoecology of the assemblage and its biogeographical relationships. The fossils are mostly preserved in channel deposits, for this was North Africa immediately after the Late Miocene "Messinian Salinity Crisis", when the earlier Tethys ocean, with its coral reefs, had retreated, the sea-level had fallen and the rivers had cut down into their beds. The connection with the Atlantic had been blocked and the Mediterranean had become a desert scattered with saline lakes.

The sea then returned with a great cataract at Gibraltar, and came close to Sahabi on the evidence of the marine mammals and microfossils. Flowing from the south was a great river with a delta, extensive saline lagoons and sabkhas. In the river were freshwater dolphins and crocodiles, and its banks were lined with forests full of small arboreal mammals. Beyond the forest was a narrow belt of

savanna, frequently afflicted with bush fires and full of animals that would seem a little strange to us today. Beyond that again was an abrupt transition to semi-arid conditions.

Although this splendid volume contains evolutionary deductions, palaeobiogeography, palaeoecology and consideration of contemporary environments, it is basically a record of what is there, described and illustrated in professional detail. It is primarily a book for specialists, and for them it fills in many of the gaps in our knowledge. It also gives the rest of us a clear picture of life in North Africa some five million years ago, when the Mediterranean had just been refilled and man was staggering to an upright stance in a world filled with a rich diversity of life. □

Derek Ager is Research Professor in the Department of Earth Sciences, University College of Swansea, Singleton Park, Swansea SA2 8PP, UK.

Plasmid practice

Simon Baumberg

Plasmids: A Practical Approach. Edited by K.G. Hardy. IRL Press: 1987. Pp.180. Hbk £25, \$45; pbk £15.50, \$28.

TRY THIS new game with your colleagues over coffee: if you had to edit a book called *Plasmids: A Practical Approach*, what would you try to include? This may imply that the contents of this volume in IRL's "Practical Approach" series are somewhat arbitrary. But the value of books such as this lies less in whether the coverage measures up to some ideal than in the quality of the contributions.

The approach in this series, maintained here, has been to combine practical detail with depth of theoretical background. All chapters score highly on both counts. Those on plasmid replication (C.M. Thomas) and incompatibility (P.L. Bergquist) have the virtue of general applicability, as would be expected for at least some contributions to a book of this title. In particular Bergquist is outstanding in marrying descriptions of modern *in vitro* techniques with updated versions of the older *in vivo* bacteriological methods from which the groups led by Datta, Meynell and Anderson derived so much knowledge 15–20 years ago: the continuing usefulness of *in vivo* work is made very plain.

The next three chapters are more specialized, but in each case there is good reason for their inclusion. P. Youngman's discussion of vectors for Tn917 transposition will be most immediately useful to *Bacillus* workers, but the approaches he describes will become increasingly useful in *Staphylococcus* and *Streptococcus*; the

topic is also one where what can be done in Gram-positive bacteria departs significantly from Gram-negative practice. A.P. Pugsley and B. Oudega provide an authoritative account of colicins and Col plasmids. Of all the chapters, this is perhaps the one whose inclusion is most debatable, but, as the editor points out, it is hard to find a recent overview elsewhere. Finally, in "Engineering an Efficient Expression System" N. Panayotatos takes the single example of pNK97, but he ably brings out considerations that would need to be taken into account in any system.

Throughout the book, techniques are described in helpful detail but often in places where one would not necessarily have thought to look for them. For instance, extraction of DNA from agarose is in Bergquist's chapter, removal of single-stranded ends with S₁ nuclease in Panayotatos's, and the use of Mudlac fusions in that by Pugsley and Oudega. Nevertheless, you can usually find what you want with the help of the index. Plasmid isolation methods for various purposes are given in all but two of the contributions: the short first chapter (by the editor), which is specifically devoted to this topic, gives the impression of being included to ensure that no potentially usable method escaped mention.

In the end, the only serious reservation I have is the inadequate coverage of conjugation; the subject is mentioned in several places, but many aspects of it are not dealt with. Otherwise this is a very useful book for research, and for undergraduate and postgraduate teaching. It will be much referred to in any laboratory or department involved in molecular genetics. □

Simon Baumberg is a Senior Lecturer in the Department of Genetics, University of Leeds, Leeds LS2 9JT, UK.

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse. I am interested in electronic ordering through the following system(s):

- ☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnTyme ☐ OCLC ILL Subsystem
☐ Other (please specify) _____
☐ I am interested in sending my order by mail.
☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name _____

Title _____

Institution/Company _____

Department _____

Address _____

City _____ State _____ Zip _____

Phone (_____) _____

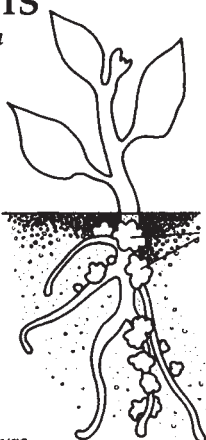
Mail to: University Microfilms International
 300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

SYMBIOSIS

An Introduction to Biological Associations

Vernon Ahmadjian and Surindar Paracer

"Can be read with profit by all professional biologists whatever their background"—*Nature*.



The authors survey the "vast field of symbiosis, both pathogenic and non-pathogenic, by the judicious selection of examples representing all types. They have succeeded admirably in this task not only by selecting interesting representative cases, but also by presenting each example in a clear and thorough way. This book brings together in one source interesting information that was previously dispersed throughout the literature... It is beautifully produced with excellent illustrations. [It] will be a valuable addition to college libraries at all levels, and general readers may also find it of interest"—*Choice*. £26.00

PREDATION

Direct and Indirect Impacts on Aquatic Communities

W. Charles Kerfoot and Andrew Sih, editors

24 essays on the effects of predation in aquatic environments, both freshwater and marine, by 32 internationally-known experts. All are previously unpublished and based on recent research. The book's theme is that the mere presence of a predator can influence interactions in competing species in many important ways, all of which have previously been nebulously termed 'indirect effects.' Written for a sizeable audience, it provides both a synthesis of latest findings and a stimulus to further research. £48.00



UNIVERSITY PRESS
OF NEW ENGLAND
Trevor Brown Associates
26 Charing Cross Road
Suite 7b, London WC2

Stringing along in search of a theory of everything

Paul K. Townsend

Superstring Theory. By M.B. Green, J.H. Schwarz and E. Witten. *Cambridge University Press: 1987. Vol.1 Introduction, pp. 469, £32.50, \$39.50. Vol. 2 Loop Amplitudes, Anomalies and Phenomenology, pp. 596, £37.50, \$49.50.*

MANY a physicist too young or too sceptical to have learned much about string theory on the occasion of its previous rise to prominence (in the context of hadron physics in the early 1970s) has bemoaned the absence of an up-to-date and comprehensive reference work on the subject. The publication of this new two-volume text has therefore been eagerly awaited. The title reflects the favoured position of supersymmetric string theories as candidates for a "theory of everything" (as the optimists put it), but all types of string theory are dealt with. Indeed, the scope of the books is enormous.

The first volume is intended to be a modern and self-contained introduction to the basic elements of string theory, covering all aspects of the classical mechanics of string theory from the point of view of the quantum mechanics of the string "worldsheet" (the first-quantized string theory). For example, the string spectrum is successively analysed by old-covariant, light-cone and modern-covariant methods, the aim being to "approach the bosonic string from many different points of view" as "each adds important ingredients to an overall understanding of string theory". The authors succeed admirably, although at the cost of some inconsistencies in the notation (about which we are warned in the preface).

The first chapter is itself an excellent review of the basic ideas of string theory, emphasizing the role of the "rubber-sheet" geometry of the two-dimensional string worldsheet, and the insights to be gained from a pictorial representation of string processes. The quantum theory of strings is also introduced here as a sum over surfaces, in analogy to Feynman's sum-over-paths approach to the quantum mechanics of particles. This modern, and geometrically appealing, approach is not given as much prominence in the second volume, where the subject of second-quantized string theory is taken up again in more detail. This is perhaps a pity, but it is in keeping with the authors' multifaceted approach to their subject.

The latter half of the first volume

introduces the complementary versions of superstring theory with, respectively, worldsheet or spacetime supersymmetry, and discusses the various types of superstring theory, including the heterotic string for which the prospects of ultimate success seem brightest. In particular, the emergence of the Yang-Mills gauge fields of the heterotic string theory is explained through the idea of toroidal compactification.

The first volume ends with a chapter on the computation of scattering amplitudes for the particles that constitute the excitation spectrum of strings. It was from the attempt to find such amplitudes, with suitable properties, that string theory was born, so this is something that one can read about in all the older reviews. Again, however, the treatment is up to date and includes a brief discussion of some very recent developments.

The second volume — *Loop Amplitudes, Anomalies and Phenomenology* — is for the more ambitious reader who wishes to find out how far string theory has gone, and how far it has yet to go, towards being a realistic theory of elementary particle physics. It effectively falls into two parts. The first deals with the quantum theory and the issues of anomalies and ultraviolet infinities at the one-loop level. The second covers the ten-dimensional supergravity equations, and the question of whether a suitable Kaluza-Klein-type solution can be found for which spacetime is effectively four-dimensional and, if so, whether a phenomenologically successful four-dimensional model can emerge. This second volume is less cohesive than the first, but such is the wealth of information contained in it that one cannot complain. Many readers will find the chapters summarizing such relatively new mathematical ideas as complex manifolds and algebraic geometry to be especially helpful.

Both volumes of *Superstring Theory* are likely to remain standard reference works for years to come; certainly, there is no serious rival at present. As the authors themselves point out, however, many topics are not covered (principally multi-loop amplitudes and string field-theory). My colleagues have also drawn my attention to a number of mistakes, dubious arguments and misprints. Such blemishes are not surprising considering the speed with which the books must have been written. This speed of delivery is perhaps partly accounted for by various passages which appear to have been taken wholesale from the authors' previous articles and reviews, without revision. One hopes that these defects will be corrected in the editions that will surely follow. □

Paul K. Townsend is a Lecturer in the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK.

Flows of researchers to and from the UK

P.M.D. Collins, G.K. Radda, J.H. Silverleaf & D.C. Smith

A new report shows that the emigration of scientists from the UK is in part balanced by immigration, and that in recent years it has been more modest in scale than is often supposed. But the matter is complicated.

MUCH has been said in the past two years about the movement of scientists and engineers out of the UK. In a climate of marked unhappiness among researchers — the President of the Royal Society stated in December 1986 that the morale of the British scientific community had “fallen to its lowest point this century”¹ — this movement has, inevitably, been labelled a ‘brain drain’ and has been taken by many commentators as symbolizing the outcome of misguided Government policies towards research. By the same token, a diminishing ‘brain drain’ would, presumably, constitute evidence that policies were right.

Evidence

This is a pity, for two reasons. First, discussion about migration of researchers has been conducted largely without the benefit of hard evidence: with rare exceptions^{2,3}, the systematic data needed to sustain an informed debate have been lacking. Secondly, migration is too complex a phenomenon to serve the symbolic function it has now acquired. To test whether migration merits the status of a ‘problem’ one must consider movement into the UK as well as movement out of the UK; one must look at how long emigrants, and immigrants, stay in their new countries, whether they return, how senior they are, how good they are, why they move; one must consider trends over time; and one must assess how easy emigrants are to replace. It is not just a question of counting heads.

Even were data on all these aspects available, the more difficult question of causal links between migration rates and Government policy would remain. Emigration could be low not because Government policies were working but, for example, because receiving countries were imposing immigration restrictions, or because potential emigrants were worried about ever being able to return to a suitable post. A high emigration rate could arise not because Government policies were failing but because a period abroad was recognized as a potentially desirable phase in one’s research career. (The latter appears to have been the case in the 1960s, when a high emigration rate coincided with what today would be considered a very high rate of real annual growth in Government expenditure on research.) So to draw conclusions about

the efficacy of Government policies for research solely on the basis of migration data is a hazardous undertaking. It is not one that we attempt here.

In this paper we report the results of a study undertaken to provide a factual basis for discussions on the flows of scientists and engineers to and from the UK⁴. The study was carried out by the Science and Engineering Policy Studies Unit (SEPSU) of the Royal Society and the Fellowship of Engineering, and looked at five broad fields of science and engineering: biochemistry, chemistry, earth science, electronic engineering and physics. We collected data by sending detailed questionnaires to a variety of research managers asking about British graduate members of their groups who had left the UK, and any members of their groups who had entered the UK, between 1975 and 1985. The questionnaires went to all relevant heads of university departments, to a representative selection of research group leaders in universities, to Research Council and other Government research institutes and to industry. The overall response rate was 77%. Some 568 questionnaires were returned.

The differences between disciplines proved to be fairly modest, so in what follows we present data for the five disciplines combined. Data on individual disciplines are given in the full report. It should be stressed, however, that our data may not be typical for all areas of science, still less for the whole of engineering, and that the picture for individual specialities may differ substantially from the overall picture for a whole discipline: the movement of a small number of people could have major consequences for a particular subdiscipline while having little effect on the discipline as a whole.

The respondents named 617 British scientists and engineers who had left jobs in the UK to work overseas, and a further 314 who left after completing higher degrees (nearly all PhDs) without having taken up employment in the UK. They also named 555 experienced scientists and engineers entering the UK, of whom 183 were British nationals returning from abroad. A further 115 immigrants came to take up their first posts after completing PhDs abroad and 15 foreign nationals stayed on after completing PhDs in the UK.

We obtained data on the size of the

groups that migrants entered or left, in order to estimate the annual migration rate. The emigration rate of British nationals from long-term posts in university departments was 0.5% per year; from research institutes and industry it was lower. Emigration from university research groups was higher, at 1.9% per year of the total membership of such groups, and emigration of recent PhDs was higher still, at about 9% per year of PhDs graduating from UK universities.

Immigration to long-term posts in university departments was 0.4% per year, and to research institutes and industry it was lower. Immigration to university research groups (including both experienced researchers and immigrants who had just obtained their PhDs) was 2.9% per year of the total membership of such groups.

These rates of migration would not in themselves generally be considered high. Even the rate for recent PhDs is considerably lower than that reported for the early 1960s. Then, about 35% of graduating PhDs emigrated each year, over half of them going to the USA⁵.

Length of stay

As we have already emphasized, however, analysis of migration is not just a matter of counting heads. One important consideration is how long migrants stay in their new countries. Here we found a clear difference between emigrants and immigrants. Among emigrants, about half the recent PhDs and nearly three-quarters of more senior leavers from universities secured long-term posts (that is, longer than three years) overseas. About 80% of the university emigrants leaving between 1975 and 1985 were known to be still abroad at the end of our survey period. In contrast, only 20% of the foreign immigrants to UK universities and research institutes took up long-term posts, and 58% of foreign immigrants were known to have left the UK again by the end of our survey. British returners were different: 77% of those returning to universities or research institutes secured long-term posts.

Nearly 60% of the emigrants in our survey went to North America. Western Europe attracted 22%. No other destination attracted more than 8%. The largest group of immigrants proved to be UK nationals returning from abroad (27%).

Western Europe provided 19% of immigrants, North America 11%, Oceania 10%, the Indian subcontinent 9%, the Far East 8% and the Middle East 7%.

Our data show a slight increase in the annual numbers both of emigrants and of immigrants during the period 1975–1985. Interestingly, the data also point to a noticeable drop in emigration rates and rise in immigration rates during the years that the University Grants Committee (UGC) 'new blood' scheme for young researchers was being introduced.

We asked respondents which factors most influenced the migrants they named. Among a considerable spread of replies, the factors most frequently identified for the university sector were career opportunities abroad (18% of all mentions), career limitations in the UK (18%), rates of pay (14%), provision of equipment and research facilities (13%) and the desire to widen experience (12%). Among non-British immigrants the pattern was different: the desire to widen experience (27%), relative status of science (15%), provision of equipment and research facilities (14%) and relative scientific vigour (12%). For British returners to universities, professional motives took a back seat to personal reasons (37%) and to the wish to see their children educated in British schools (17%).

Quality

Many respondents emphasized the need to pay attention to quality as well as quantity. Rigorous assessment of quality is obviously difficult to achieve within the context of a study of this sort. We asked respondents how many of the migrants they mentioned were 'outstanding': 40% of emigrants and 30% of immigrants were so described. Senior migrants were more likely than junior migrants to be described as 'outstanding'.

We asked other subjective questions. Two-thirds of respondents indicated that emigrants were difficult to replace with people of comparable experience and expertise, especially if they were reasonably senior. Opinions about employment opportunities for new PhDs in the UK were divided: in general, while short-term prospects were adequate, long-term career prospects were held to be rather poorer. The impact of migration on UK research was reckoned to be more severe in universities and research institutes, where 73% thought it was having an adverse effect, than in industry, where 45% thought so.

The SEPSU report is the only source of recent detailed data on emigrants and immigrants, but some relevant data are also to be found elsewhere. The Universities Statistical Record (USR), for example, has information on the numbers of staff leaving UK universities. The annual number of staff in USR subject

categories III to V (engineering and technology, agriculture, and biological and physical sciences) who went to jobs overseas rose from 350 in 1974 to 447 in 1977 and then declined to 290 in 1984. Of these emigrants, the numbers who left 'wholly university financed' posts (roughly equivalent to long-term posts in university departments) fell, while the numbers who left posts financed by non-university sources (roughly equivalent to the research group leavers in the SEPSU survey) rose. Generally, the USR data corroborate the SEPSU evidence that the rate of emigration is, numerically, fairly low and has not grown rapidly in recent years.

The National Science Foundation (NSF) analyses data on scientists and engineers given permanent immigration status in the USA. This is the source of the much-quoted figure of 1,000 scientists and engineers from the UK going annually to the USA. In 1985, in fact, 979 scientists and engineers whose last country of residence had been the UK gained permanent immigration status in the USA; many may have actually migrated to the USA before that date. Of the 979, only 865 were UK citizens. Of these, 134 were natural scientists, 597 were engineers and 121 were mathematicians or computer scientists. The annual figure for natural scientists has not changed greatly since the early 1970s, but that for engineers and mathematicians has more than doubled in the same period.

Data on the geographical distribution of Fellows of the Royal Society have been published already¹. In 1960, 2.8% of the ordinary Fellowship was living in the USA, compared with 8.1% in 1986. The figures for Fellows living in the USA at the time of their election are higher: 3.9% and 13.2%. An increasing number of the most talented British scientists have been among those who have left the UK.

The available evidence thus suggests that the outflow of British scientists from the UK has been at a fairly modest level in recent years and, at the end of 1985, was not increasing rapidly. There is some indication that, for engineers, the picture may be generally more serious. The inflow to the UK of scientists and engineers, whether foreign nationals or returning British citizens, has largely balanced in numerical terms the outflow of experienced researchers, though not of recent PhDs. However, because emigrants tend to go to long-term posts and immigrants to short-term posts, the net loss of scientific capacity from the UK is greater than simple head-counting would imply. Moreover, it appears that emigration is increasingly affecting the most talented researchers, not least because the receiving countries are now applying stricter criteria for immigration permits.

Scientists who have just obtained their

PhDs and are looking for their first substantive posts are the most likely to emigrate, scientists on short-term contracts less likely, and those in long-term positions least likely to do so. The importance of providing long-term career prospects for research scientists and engineers emerges clearly from the SEPSU survey, as does a widespread feeling that such prospects are currently pretty poor in the UK. The introduction of the UGC 'new blood' scheme coincided with a decrease in the net annual outflow of researchers, and a number of respondents argued for the retention of this or a similar scheme.

Industry

There are signs that, for the university system, the movement of staff in some disciplines to industry and commerce within the UK is likely to become more serious than the movement of staff overseas. That industry appears to be less concerned about migration than the universities or research institutes reflects its ability to compete more successfully for the limited pool of available researchers. The difficulties that universities are now experiencing in attracting good new recruits to an academic career in certain fields mean that the movement of staff to overseas posts, even on the scale reported in the SEPSU survey, lessens the capacity of the research system in the UK to continue to meet the needs of the nation in its wealth-creating activities.

The widespread concern expressed in recent months about the migration of scientists and engineers is part of a more general concern about the well-being of the British research system as a whole. Migration is only one of a range of pertinent factors, though its emotive nature has given it prominence in public discussion. We have shown that migration is a complex phenomenon and that its scale is more modest than some might have anticipated. We have also shown that our data provide no grounds for complacency. □

1. Address of the President, Sir George Porter, Given at the Anniversary Meeting on 1 December 1986 *Science and Public Affairs* 2, 4–8 (1987).
2. *The Brain Drain: Summary of Findings of Enquiry* (Advisory Board for the Research Councils, London, 1985).
3. Pearson, R & Parsons D. *The Biotechnology Brain Drain* (Science and Engineering Research Council, Swindon/Institute of Manpower Studies, Brighton, 1983).
4. *The Migration of Scientists and Engineers To and From the UK SEPSU Policy Study No 1*. (Royal Society/Fellowship of Engineering, London, 1987).
5. *Emigration of Scientists From the UK* (Royal Society, London, 1963).
6. Smith, D.C., Collins, P.M.D., Hicks, D.M., & Wyatt, S. *Nature* 323, 681–684 (1986).

Peter Collins is Head of the Science and Engineering Policy Studies Unit of the Royal Society and the Fellowship of Engineering. George Radda FRS was Chairman of the Task Group appointed to oversee the study. Jane Silverleaf, who carried out the research reported here, is a member of the Unit. Sir David Smith, Biological Secretary of the Royal Society, is Chairman of the SEPSU Steering Group.

Information-based complexity

Edward W. Packel

Department of Mathematics and Computer Studies, Lake Forest College, Lake Forest, Illinois 60045, USA

J. F. Traub

Computer Science Department, Princeton University, Princeton, New Jersey 08544, USA

Information-based complexity seeks to develop general results about the intrinsic difficulty of solving problems where available information is partial or approximate and to apply these results to specific problems. This allows one to determine what is meant by an optimal algorithm in many practical situations, and offers a variety of interesting and sometimes surprising theoretical results.

MANY problems arising in the sciences have the characteristic that information relevant to their solution is either partial or contaminated. For example, as a digital computer can manipulate only a finite set of numbers, any problem whose domain of possible problem elements is infinite will of necessity have only partial information. For most problems with partial or contaminated information, only approximate solutions are possible. Given the ubiquity of problems for which information is not fully available, imprecise because of computational limitations or is purposely discarded for the sake of simplified solution, a general approach to the approximate solution of such problems is clearly desirable.

Whether or not full information is available, mathematicians and computer scientists have recently paid close attention to the study of the inherent complexity of problems as defined in terms of the computational resources required for their solution. The resulting field of research, computational complexity, has provided a wealth of exciting results of both a theoretical and practical nature. From this has sprung the field of information-based complexity, an expanding area of research concerned with the intrinsic difficulty of solving problems approximately on the basis of information that is partial, contaminated or priced. Detailed surveys can be found in refs 1–3. In what follows, we summarize some of the key aspects of information-based complexity, its relationship to other areas of inquiry, its mathematical formulation, some key areas of application and some surprising theoretical results.

Thermodynamics of computation

Classical thermodynamics has two general characteristics. First, it sets intrinsic limits on what heat engines can do. Second, it provides benchmarks for any particular engine, whence the notion of efficiency. Computational complexity is analogous. It sets intrinsic limits on what problems can be solved. Thus the computational complexity of a problem is measured by the minimal computational resources (time or memory, for example) required for its solution. Problems that cannot be solved because practical limitations dictate that the requisite computational resources can never be achieved are said to be intractable.

Readers should please note that the term computational complexity has served double duty both as the name of a field and as a crucial invariant characteristic of the problem.

Having provided a benchmark for the intrinsic difficulty of a problem, techniques of computational complexity can be applied to any algorithm that solves a particular problem to tell how well it measures up. This leads to the important notion of an optimal algorithm, one that solves a problem with a 'cost' commensurate with its computational complexity.

Two kinds of complexity

Computational complexity is of two kinds. On the one hand, there is that kind of complexity typified by the travelling salesperson problem (henceforth TSP), a discrete problem that conveys the flavour of what is known as combinatorial complexity. The second branch of computational complexity, information-based complexity, is exemplified by the continuous problem of computing the definite integral for a class of functions of one variable.

The TSP is formulated as follows. One is given a set of n cities together with the distances between all pairs of cities. A trip must be made that visits each city exactly once and returns to the starting point. The problem is to find a tour, essentially an ordering of the n cities, that minimizes the total distance travelled. This aptly named problem turns out to be an abstraction of many scheduling and layout problems in the real world.

Two features of this problem stand out. The first relates to the smallest number of computational steps (additions, multiplications, comparisons, for example) required to solve the TSP as a function of the number of cities; formulated carefully, this number defines the computational complexity of TSP. Remarkably, even the general order of magnitude of the computational complexity of this problem is at present unknown. Despite intensive study, the best algorithms for finding an optimal tour of n cities require an exponential number of steps. We call computational complexity which is exponential in the problem size n intractable⁴.

A second characteristic of TSP is that the available information is complete, so that, in principle, the problem can be solved exactly. Given the intercity distances, the determination of the distance travelled for a minimal tour becomes a matter of systematic computation. Estimating the number of steps required by an algorithm to find an exact solution (thus fixing an upper bound on the complexity of the problem) and estimating a minimum number of steps over all such algorithms (yielding a corresponding lower bound) often requires difficult combinatorial arguments. Thus the branch of computational complexity concerned with problems with complete information and having exact solutions is called combinatorial complexity.

The early emphasis in computational complexity is usually associated with discrete problems such as TSP, bin packing and testing for primality, but there is also a substantial literature, starting with the work of Sard⁵, Nikolskij⁶ and Kiefer⁷, on the complexity of continuous problems. More recently, mathematicians have been interested in the complexity of continuous problems, sparked by the work of Shub and Smale^{8–10}.

Consider, in that context, the problem of computing definite integrals on a finite interval. More formally, let $C[0, 1]$ denote

the space of continuous real-valued functions on the closed interval $[0, 1]$ endowed with the sup norm,

$$\|f\| = \sup_{t \in [0, 1]} \{|f(t)|\},$$

and consider the subspace F defined by

$$F = \{f \in C[0, 1]: |f'(t)| \leq 1 \text{ for all but finitely many } t \in [0, 1]\}.$$

The problem we wish to solve is defined by $S: F \rightarrow \mathbb{R}$, where $S(f) = \int_0^1 f(t) dt$ and $\mathbb{R}$ is this field of real numbers. The class of integrands must be restricted for the computational complexity to be finite under a worst-case-error criterion. In this example we require that the absolute value of the derivative of $f(t)$ be less than 1 almost everywhere.

In the friendly world of elementary calculus, one is led to expect exact solutions to this problem, but in most real situations, we cannot compute definite integrals by obtaining the anti-derivative symbolically. Instead, we must resort to numerical integration, which takes this form: the integrand f is evaluated at certain points t_i , $i = 1, 2, \dots, n$ and the numbers t_i and the function values $f(t_i)$ are then combined to provide an estimate of the integral.

In general, we cannot provide a computer with the actual function to be integrated. Having selected the n points t_i to work with, we are permitted to use only the values $f(t_i)$ as the available information in attempting to compute the definite integral $S(f)$. Because there are infinitely many functions f in F that share the same information, we cannot recover the 'true' integrand. Thus, the problem has an intrinsic uncertainty independent of the numerical integration algorithm used. In contrast with the complete informational setting of TSP, this problem has information that is partial.

This example points to the general problem of information-based complexity, another branch of computational complexity. What does computational complexity mean for a problem with only partial information? If we are willing to tolerate an uncertainty of ε in the accuracy of our estimate of the definite integral, we would like to determine the smallest amount of computational resources needed to achieve that accuracy. With fixed uncertainty ε , what can we say about intractability in the presence of partial information?

Information-based complexity

We are, in general, concerned with computational complexity of problems for which the information is partial, contaminated and priced. Three concepts provide the foundation for developing and applying techniques of information-based complexity: a problem formulation, specification of the information and a model of computation.

The problem formulation states what we want to approximate, for which problem elements we seek this approximation and what error criterion we plan to use. Formalizing these ideas, if F is a set and G a real normed linear space, the problem we would ideally like to solve is given by a mapping $S: F \rightarrow G$ called the solution operator. Given $\varepsilon \geq 0$, we wish to compute for each $f \in F$ an ε -approximation $x(f) \in G$ for $S(f)$. Under a worst-case error-criterion¹¹ we would require that $x(f)$ satisfy $\|S(f) - x(f)\| \leq \varepsilon$ for every problem element f . An average case error criterion¹² assumes a probability measure on the space F and would require that an appropriate average error between $S(f)$ and $x(f)$ not exceed ε . Other error criteria are also possible¹³⁻¹⁶, but we shall focus on the traditional worst case and on the increasingly important average case.

To approximate $S(f)$, we need to know something about f . Formally, knowledge about f is obtained through an information operator $N: F \rightarrow \mathbb{R}^n$, where $\mathbb{R}^n$ is the n -dimensional field of real numbers. The requirement that N have finite dimensional range (n) is a concession to the fact that practical algorithms process only a finite amount of information; as a consequence, there

will in general be many different problem elements f having the same information $N(f)$. One way to view this situation¹⁷ is that N decomposes the space of problem elements into equivalence classes in each of which all the members of a class have the same information. The challenge is then to approximate $S(f)$ knowing only the equivalence class to which f belongs.

Any analysis of complexity requires a model of computation which specifies what operations are permitted and how much each operation 'costs'. We use what is referred to as the real-number model of computation, in which we assume:

- real numbers of infinite precision are used;
- each information operation costs c units;
- comparisons and arithmetical operations such as addition and scalar multiplication in G are performed exactly and cost 1 unit each.

More detailed models of computation are possible, but we use this simplified model to avoid being sidetracked by such admittedly important issues as round-off, machine specificity and numerical stability.

Recalling the integration example discussed earlier, we note that the solution operator $S: F \rightarrow \mathbb{R}$ is defined by the definite integral. The information operator $N: F \rightarrow \mathbb{R}^n$ is defined, for a preselected set of n points t_i ($0 \leq t_1 < t_2 < \dots < t_n \leq 1$), by $N(f) = [f(t_1), f(t_2), \dots, f(t_n)]$. Formally, the TSP can also be viewed in the framework we have described by letting F denote the set of networks (cities together with intercity distances) and defining $S(f)$ to be a minimal tour for the network f . As information in this case is complete, $N: F \rightarrow F$ will be the identity operator. We hasten to add that researchers in combinatorial complexity do not tend to think of their subject in this way. The dichotomy that we described for computational complexity in the previous section can be summarized as in the box below.

Combinatorial complexity	Information-based complexity
• Problems are discrete • Information is complete, exact and free • Given f compute $S(f)$	• Problems are continuous • Information is partial, contaminated and priced • Given $N(f)$ approximate $S(f)$

For any $f \in F$, the solution value $S(f)$ will be approximated by an algorithm $\Phi: N(F) \rightarrow G$. As the notation suggests, the approximation $\Phi(f)$ provided by an algorithm must be based on the information $N(f)$ rather than on the (unknown) problem element f . Thus, in the integration example, one possible algorithm could be defined for information

$$N(f) = [f(t_1), f(t_2), \dots, f(t_n)]$$

by

$$\Phi(N(f)) = \sum_{i=1}^n f(t_i)[t_i - t_{i-1}] \quad (t_0 = 0),$$

which is a simple rectangular integration method.

The radius of information

An important concept in information-based complexity is that of the radius of information, $r(N)$, which measures the error or uncertainty intrinsic to a problem S with information N . In the worst case, this is defined as follows. Let $r(N, f)$ denote, for each $f \in F$, the radius in the normed space G of the set $S([f])$, where $[f]$ denotes the equivalence class of problem elements having the same information as f . Then

$$r(N) = \sup_{f \in F} \{r(N, f)\}.$$

Some of the key ideas we have developed about information-based complexity are summarized in Fig. 1. We are given information $N(f)$ about problem elements f . In a worst case setting, a maximally 'unfriendly' problem element f is anticipated, resulting in the largest possible radius among the sets $S[f]$. The radius of information arising from this process gives us a lower

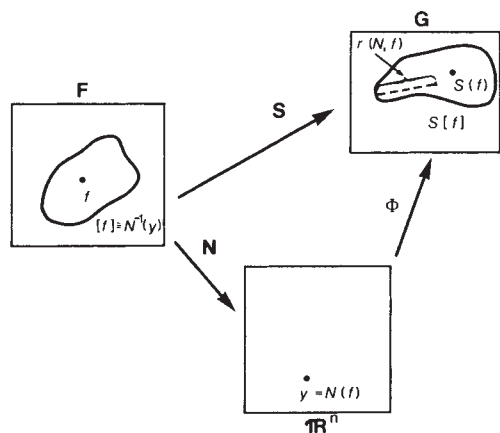


Fig. 1 Solution operator S , information operator N and algorithm Φ .

bound on how closely any algorithm Φ can approximate the solution $S(f)$ by $\Phi(N(f))$.

To quantify this notion of lower bound, we let

$$e(\Phi, N) = \sup_{f \in F} \{\|S(f) - \Phi(N(f))\|\}$$

denote the worst-case error of the algorithm $\Phi: N(F) \rightarrow G$. The following key result is readily established¹¹.

Theorem. $r(N) = \inf \{e(\Phi, N) : \Phi \text{ an algorithm using information } N\}$.

Thus the radius of information gives us a sharp lower bound on the error obtained in approximating S by any algorithm using the given information. It is natural, then, to define Φ to be an optimal error algorithm for the problem S with information N if $e(\Phi, N) = r(N)$. We note that definitions and a theorem analogous to these can also be developed in an average case setting^{2,12}.

Complexity and optimality

Given an algorithm Φ and a model of computation, it is just a matter of bookkeeping to determine the cost of Φ . Given $f \in F$, let $\text{cost}(N, f)$ denote the information cost of computing $N(f)$ and let $\text{cost}(\Phi, N(f))$ denote the combinatory cost of computing $\Phi(N(f))$. We can now define our fundamental notion of complexity in the worst case setting. Given an error tolerance $\varepsilon \geq 0$ for a problem $S: F \rightarrow G$, we define the ε -complexity as the minimal worst case cost of computing an ε -approximation,

$$\text{comp}(\varepsilon) = \inf_{\Phi, N: e(\Phi, N) \leq \varepsilon} \sup_{f \in F} \{\text{cost}(N, f) + \text{cost}(\Phi, N(f))\}.$$

An information operator N and an algorithm Φ using N for which this infimum is obtained are defined to be optimal ε -complexity information and an optimal ε -complexity algorithm respectively. Since ε is regarded as fixed, we omit ' ε -complexity' in future references to these important concepts.

To illustrate these ideas, we return to our integration example. In Fig. 2, it can be seen that for the function evaluation information pictured, the two saw-tooth functions severally represented by dotted and dashed lines are evaluated at the three points t_1 , t_2 and t_3 and have identical information N . The quantity $r(N, f)$ is the area between either bounding sawtooth function and the piecewise linear function that 'bisects' the sawtooth functions. The largest such area, and hence the radius of information $r(N)$, results from functions with zeros at the fixed evaluation points.

If we are free to choose n evaluation points $t_1, \dots, t_n$ ourselves, it can be shown that the best choice for minimizing $r(N)$ is to regard $[0, 1]$ as a circle by making 0 and 1 identical and to choose the t_i to be 'equally spaced'. For this choice of N , we have $r(N) = 1/4n$. Given an error tolerance $\varepsilon > 0$, the number of equally spaced evaluation points must satisfy $n \geq 1/4\varepsilon$ to ensure $r(N) \leq \varepsilon$. It follows that optimal information consists of

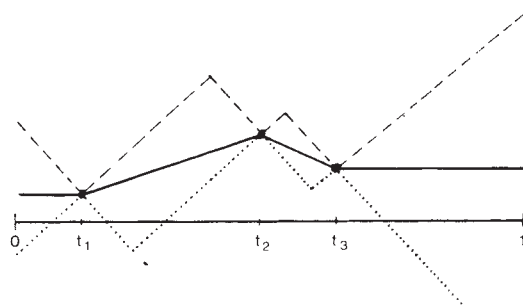


Fig. 2 Bounds and centre for functions $g \in F$ with $g(t_i) = f(t_i)$, $i = 1, 2, 3$.

n equally spaced points, where $n = \lceil (1/4\varepsilon) \rceil$. (The quantity $\lceil x \rceil$ denotes the smallest integer at least as large as x .) A particularly simple optimal algorithm then results by defining $\Phi(N(f)) = 1/n \sum_{i=1}^n f(t_i)$, where the t_i are the equally spaced evaluation points. It can be shown from our model of computation that, to within at most one unit, $\text{comp}(\varepsilon) = (c+1)\lceil (1/4\varepsilon) \rceil$.

The linear optimal algorithm that results above is essentially trapezoidal integration, not the most honoured of quadrature methods in terms of its speed of convergence as n increases. It is, however, optimal in the worst-case sense that we have described. The fact that our linear integration problem has a linear optimal error algorithm is more than fortuitous. There are a variety of general results about existence of linear optimal error algorithms for linear problems as well as some interesting counterexamples^{18-21,53}.

Application areas

Instead of hand-crafting algorithms, often on the basis of *ad hoc* criteria, information-based complexity can provide benchmarks and powerful tools for mechanizing the creation of optimal or nearly optimal algorithms. To be genuinely useful, algorithms must also enjoy properties such as ease of implementation and numerical stability. These properties can be taken into account in the general theory.

Although our illustrative example is a rather simple integration problem, we note that techniques of information-based complexity have the potential to be applied to virtually any well-defined problem where complete information is not available. We itemize below, with references, some of the problems for which methods of information-based complexity are being used by researchers. The problems range from specific applications to general mathematical topics:

- Human and robotic vision²²⁻²⁴
- Signal processing²⁵
- Prediction and estimation^{26,27}
- Synchronization of clocks in distributed systems²⁸⁻³⁰
- Approximation and its applications³¹
- Nonlinear constrained optimization³²
- Remote sensing and ill-posed problems^{33,34}
- Multivariate integration³⁵⁻³⁷
- Systems of nonlinear equations^{8-10,38-40}
- Large eigenvalue problems⁴¹
- Ordinary differential equations^{42,43}
- Partial differential equations and integral equations⁴⁴
- Large linear systems⁴⁵

We refer the reader to the recent survey by Woźniakowski² where descriptions and more complete references are given for current work in many of the above areas.

Intractable problems

We now return to the issue of intractability. We noted earlier that, in the realm of combinatorial complexity, TSP is believed, but not proven, to be an intractable problem. In fact TSP, along

with many other much studied problems, is centrally important to theoretical computer science as an NP-complete problem⁴. Here we consider the matter of intractable problems in information-based complexity.

Given an error tolerance $\varepsilon > 0$, it has been shown for many problems in mathematical analysis that the time complexity (number of operations) is proportional to $(1/\varepsilon)^{d/r}$ where d is the dimension of the problem and r measures the smoothness^{2,46}. Note that, for fixed smoothness, the complexity is exponential in d . Among the problems with complexity proportional to $(1/\varepsilon)^{d/r}$ are multivariate integration³⁵⁻³⁷, multivariate approximation⁴⁷, and systems of nonlinear equations⁴⁰.

A numerical example illustrates the implications of this complexity function. If one wants eight-place accuracy in computing triple integrals of functions that are once differentiable, then $\varepsilon = 10^{-8}$, $d = 3$, $r = 1$, and $(1/\varepsilon)^{d/r} = 10^{24}$. Even if we generously assume the existence of a sequential computer that performs 10^{10} operations per second, the computation will take of the order of 10^{14} seconds or some 100,000 years.

By comparing what we have said about TSP with multivariate problems of the type discussed here, we note two important general differences. Exponential complexity for TSP (and many other problems) is currently only conjectured, but exponential complexity has been established for the indicated problems in information-based complexity. Secondly, the role of the problem size n in combinatorial complexity is, in information-based complexity, often taken by the dimension d . We observe that problems of high dimensionality occur in such diverse disciplines as physics and nonlinear economic modelling.

These conclusions are based on complexity considerations, meaning that they are intrinsic to the problem we are trying to solve. No algorithm, no matter how clever, can sidestep these computational barriers.

Can intractable problems be solved?

In spite of the seemingly absolute nature of these conclusions on intractability, there are several ways in which their consequences might be mitigated.

Physicists have long known about the power of randomness. Classical Monte Carlo methods, for example, may be regarded as random processes for making the cost of certain computations independent of dimension. When randomness is used, one must be willing to settle for probabilistic rather than worst-case assurances about error. Researchers in information-based complexity are now asking basic questions about the power and limitations of randomness. What problems can be solved more easily by using randomness? How much can the complexity of such problems be lowered? How much can their uncertainty be reduced? Are there problems for which randomness can be shown to be of no help⁵⁴?

The intractability results discussed in the last section are based upon a worst case setting for our problem formulation. If one is willing to live with the assurance that the uncertainty is no greater than ε on average, can that make tractable a problem that is intractable in the worst case? Average case analysis of problems and algorithms is an increasingly important area of study in theoretical computer science. Much current work in this area is also taking place in information-based complexity. The investigation of the extent to which complexity is lowered in the average case setting is an important part of this work^{2,3,12,48,49}.

Another approach to measuring uncertainty and complexity uses a probabilistic setting, in which one allows uncertainty to exceed ε , but only with a small predefined probability. Questions of error and intractability in this setting are also the object of current research^{14,15,48}.

Parallel computation

In the integration example discussed earlier and in many other important problems the information operators used have the

form $N(f) = [L_1(f), L_2(f), \dots, L_n(f)]$, for a fixed number n of linear functionals $L_i: F \rightarrow \mathbb{R}$ (the problem domain F is assumed to be a linear space). Each component of such information can be computed independently so that $N(f)$ can be separated into n 'parallel' computations. Information with the above property is called nonadaptive information and is of considerable importance in facilitating parallel computation. Indeed, if n nonadaptive evaluations are required to achieve the desired accuracy we may use n processors to perform the desired evaluations simultaneously. This will essentially lower the information cost by a factor of n .

If, on the other hand, the choice of the L_i depends on the previously computed information or if the number of evaluations n varies with f , the information is called adaptive and will tend to be unsuitable for parallel computation. A question that naturally arises is the extent to which nonadaptive information can be computationally as powerful as adaptive information. This question can be carefully formulated in the context of information-based complexity. It has been shown^{11,50,51} that for linear problems (S and N are linear operators), nonadaptive information is essentially as powerful as adaptive information. This result is surprisingly robust, holding in both worst-case and average-case settings^{48,51}.

Thermodynamics and information theory

We began this article by observing that computational complexity and thermodynamics share two characteristics. We conclude by discussing how information-based complexity relates to the notion of entropy, a term whose origins come from thermodynamics. Our interest here is in a more recent use of the term entropy and its by-product, mutual information, as defined in classical information theory.

Concepts to measure uncertainty, disorder and information content have arisen in a variety of scientific disciplines. Entropy, introduced by Clausius in thermodynamics, was first used in a probabilistic setting by Boltzmann in his treatment of statistical mechanics. In statistics, the variance and R. A. Fisher's technical definition of information are noteworthy. In mathematics, Kolmogorov used the term entropy to define a measure of 'capacity' in abstract metric spaces. Finally, Shannon introduced the logarithmic definitions of information and its negative, entropy, as the basis for information theory.

Information-based complexity has connections with several of the concepts discussed above. Thus, in a worst-case setting, there are connections between ε -complexity and Kolmogorov entropy¹¹. Working in an average case setting, there is a class of problems for which the radius information $r^{\text{ave}}(N_n)$ can be expressed as the standard deviation of a related distribution⁴⁸.

Finally, one can define a value of information $V(N) = \log_2(r^{\text{ave}}(0)/r^{\text{ave}}(N))$, where the notation indicates an average-case setting and $r^{\text{ave}}(0)$ denotes the radius computed with no information available. It can be shown that $V(N)$ provides interesting connections and contrasts with the corresponding average mutual information, an entropy-based concept from information theory. The two values agree for an important class of problems. In a variety of other situations, however, the average mutual information is not well-defined or the two values disagree. The value of information $V(N)$ is virtually always defined and has a natural interpretation in terms of 'bits of accuracy gained'. Thus for problems with incomplete information, where information provides a mechanism for refining what we know about the space of problem elements, information-based complexity provides a new approach to measuring the value of information.

We thank the National Science Foundation for partial support.

1. Traub, J. F. & Woźniakowski, H. in *Advances in Computers* 23, (ed. Yovits, M.) 35-92 (Academic, New York, 1984).
2. Woźniakowski, H. *A. Rev. Comput. Sci.* 1, 319-380 (1986).
3. Packel, E. W. & Woźniakowski, H. *Bull. Am. Math. Soc.* (in the press).

4. Garey, M. R. & Johnson, D. S. *Computers and Intractability* (Freeman, San Francisco, 1979).
5. Sard, A. *Am. J. Math.* **71**, 80-91 (1949).
6. Nikolskij, S. M. *Usp. Math. Nauk* **5**, 165-177 (1950).
7. Kiefer, J. *Proc. Am. Math. Soc.* **4**, 502-505 (1953).
8. Shub, M. & Smale, S. *Annales Scientifiques de L'Ecole Normale Supérieure 4 serie*, **18** (1985).
9. Shub, M. & Smale, S. *J. Complexity* **2**, 2-11 (1986).
10. Smale, S. *Bull. Am. Math. Soc.* **13**, 87-121 (1985).
11. Traub, J. F. & Woźniakowski, H. *A General Theory of Optimal Algorithms* (Academic, New York, 1980).
12. Wasilkowski, G. W. *J. Complexity* **1**, 107-117 (1985).
13. Wasilkowski, G. W. & Woźniakowski, H. *Comp. Sci. Department Report* (Columbia University, 1984).
14. Woźniakowski, H. *J. Complexity* **2**, 255-269 (1986).
15. Woźniakowski, H. in *Proceedings of Optimal Algorithms* (Bulgarian Academy of Science, Sofia, 1986).
16. Wasilkowski, G. W. *Comput. Sci. Dept Rep.* (Columbia University, 1987).
17. Bartholdi, J. J. & Ford, J. *Physics Today* **38**, 75-76 (June, 1985).
18. Smolyak, S. A. Dissertation, Moscow State Univ. (1965).
19. Osipenko, K. Yu. *Mat. Zametki* **19**, 29-40 (1976). (English Translation *Math. Notes* **19**, 17-23 (1976).)
20. Packel, E. W. *Aequationes Math.* **31**, 18-25 (1986).
21. Werschulz, A. G. & Woźniakowski, H. *Aequationes Math.* **31**, 202-211 (1986).
22. Poggio, T., Torre, V. & Koch, C. *Nature* **317**, 314-319 (1985).
23. Lee, D. *J. Complexity* **1**, 138-146 (1985).
24. Lee, D. thesis, Columbia University (1985).
25. Kowalski, M. A. *J. Complexity* **2**, 239-254 (1986).
26. Milanese, M. & Tempo, R. *IEEE Trans. Autom. Contr.* **AC 30**, 730-738 (1985).
27. Milanese, M., Tempo, R., & Vicino, A. *J. Complexity* **2**, 78-94 (1986).
28. Halpern, J. Y., Simons, B. B., Strong, H. R., & Dolev, D. in *Proceedings of the 3rd ACM Conference on Principles of Distributed Computing*, 89-102 (1984).
29. Lynch, N. A. & Lundelius, J. *Information and Control* **62**, 190-204 (1985).
30. Wasilkowski, G. W. *Comput. Sci. Dept Rep.* (Columbia University, 1986).
31. Micchelli, C. A. & Rivlin, T. J. in *Optimal Estimation in Approximation Theory* (eds Micchelli, C. A. & Rivlin, T. J.) 1-54 (Plenum, New York, 1977).
32. Nemirovsky, A. S. & Yudin, D. B. *Problem Complexity and Method Efficiency in Optimization* (Wiley, New York, 1983).
33. Twomey, S. *Introduction to Mathematics of Inversion in Remote Sensing and Indirect Measurements, Developments in Geomathematics* **3** (Elsevier, Amsterdam, 1977).
34. Werschulz, A. G. *Comput. Sci. Dept Rep.* (Columbia University, 1986).
35. Bakhvalov, N. S. *U.S.S.R. Comput. Math. Phys.* **68-82** (1961).
36. Babenko, V. F. *Math. Notes* **19**, 187-193 (1976).
37. Sukharev, A. G. *Sov. Math. Dokl.* **20**, 472-475 (1979).
38. Schonhage, A. *Report* (University of Tübingen, 1982).
39. Renegar, J. *Math. Dept. Rep.* (Colorado State Univ., 1985).
40. Sikorski, K. *J. Complexity* **1**, 197-209 (1985).
41. Kuczyński, J. *J. Complexity* **2**, 131-162 (1986).
42. Kaciewicz, B. Z. *BIT* **23**, 217-230 (1983).
43. Kaciewicz, B. Z. *Numer. Math.* **43**, 93-104 (1984).
44. Werschulz, A. G. *J. Complexity* **1**, 232-255 (1985).
45. Traub, J. F. & Woźniakowski, H. *Journal of the ACM* **31**, 545-559 (1984).
46. Werschulz, A. G. *J. Integral Equations* **9**, 213-241 (1985).
47. Babenko, K. I. in *Theoretical Background and Constructing of Computational Algorithms for Mathematical-Physical Problems* (ed. Babenko, K. I.), (Nauka, Moscow, 1979).
48. Lee, D. & Wasilkowski, G. W. *J. Complexity* **2**, 12-43 (1986).
49. Papageorgiou, A. & Wasilkowski, G. W. *Comput. Sci. Dept Rep.* (Columbia University, 1986).
50. Gal, S. & Micchelli, C. A. *Appl. Anal.* **10**, 105-120 (1980).
51. Wasilkowski, G. W. *J. Complexity* **2**, 209-228 (1986).
52. Packel, E. W., Traub, J. F. & Woźniakowski, H. *Comp. Sci. Department Report* (Columbia University, 1987).
53. Packel, E. W. *SIAM Rev.* (in the press).
54. Traub, J. F., Wasilkowski, G. W. & Woźniakowski, H. *Information-Based Complexity* (Academic, New York, in the press).

ARTICLES

Transgenic plants protected from insect attack

Mark Vaeck, Arlette Reynaerts, Herman Höfte, Stefan Jansens, Marc De Beuckeleer, Caroline Dean*, Marc Zabeau, Marc Van Montagu & Jan Leemans

Plant Genetic Systems NV, Jozef Plateastraat 22, B-9000 Gent, Belgium

The Gram-positive bacterium Bacillus thuringiensis produces proteins which are specifically toxic to a variety of insect species. Modified genes have been derived from bt2, a toxin gene cloned from one Bacillus strain. Transgenic tobacco plants expressing these genes synthesize insecticidal proteins which protect them from feeding damage by larvae of the tobacco hornworm.

MODERN agriculture uses a wide variety of insecticides to control insect damage. Most of them are chemically synthesized. Notable exceptions are the insect toxins produced by *Bacillus thuringiensis*: spore preparations of this Gram-positive bacterium have been used for more than 20 years as a biological insecticide¹. The insecticidal activity resides in crystalline inclusion bodies produced during sporulation of the bacteria, which are composed of proteins (termed delta endotoxins) specifically toxic against a variety of insects. Different strains of *B. thuringiensis* differ in their spectra of insecticidal activity. Most are active against Lepidoptera, but some strains specific to Diptera^{2,3} and Coleoptera^{4,5} have been identified. The crystals dissolve in the alkaline conditions of the insect midgut and release proteins of relative molecular mass 65,000-160,000 (M_r 65K-160K)^{2,5,6} which are proteolytically processed by midgut proteases to yield smaller toxic fragments⁷. *B. thuringiensis* insect toxins are highly specific, in that they are not toxic to other organisms. Hence, they are safe insecticides and present an interesting alternative to chemical control agents. Their commercial use however is limited by high production costs and the instability of the crystal proteins when exposed in the field.

We have used *Agrobacterium*-mediated T-DNA transfer⁸ to express chimaeric *B. thuringiensis* toxin genes in tobacco plants with the objective of protecting the plants from insect attack.

* Permanent address: Advanced Genetic Sciences Inc., San Pablo Avenue, 6701, Oakland, California 94608, USA.

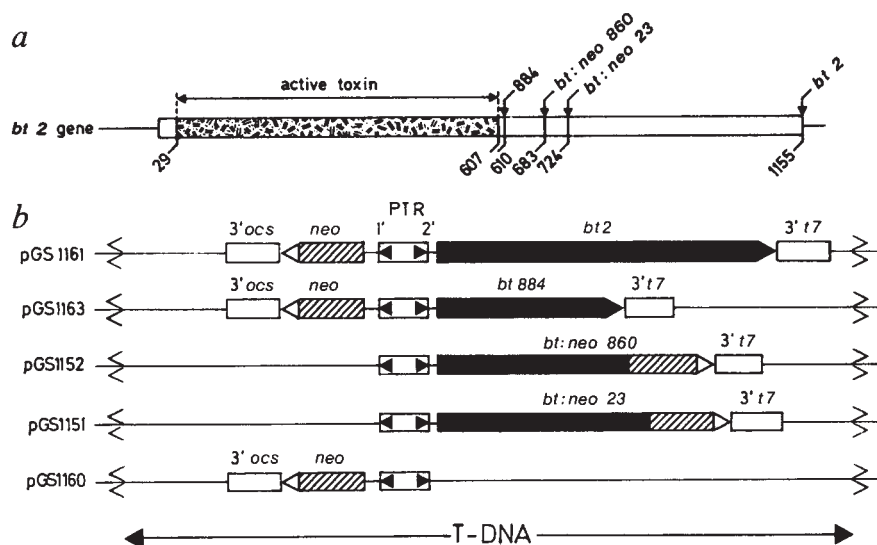
We show here that a defence mechanism against phytophagous insects can be devised by genetic engineering.

Modified *Bacillus* toxins

We have reported⁹ the cloning of the *bt2* gene from *B. thuringiensis* strain berliner 1715 and the characterization of the recombinant polypeptide expressed in *Escherichia coli*. This protein, termed Bt2, is 1,155 amino acids long and is a potent toxin to several lepidopteran larvae, such as those of *Manduca sexta*, a pest of tobacco. Bt2 is a protoxin and generates a smaller polypeptide of M_r 60K which retains full toxic activity⁹. The smallest fragment of Bt2 that is still fully toxic has been mapped in the NH₂-terminal half of the protein, between amino-acid positions 29 and 607 (ref. 9) (and see Fig. 1a).

In plant transformation experiments, we used chimaeric genes containing the entire coding sequence of *bt2* as well as truncated genes. A diagram of the chimaeric genes is shown in Fig. 1b. Some of our T-DNA constructs include a chimaeric neomycin phosphotransferase gene (*neo*) as a marker selectable in plants¹⁰. Others carry translational fusions between fragments of *bt2* and the *neo* gene. Fusions to the 5' end of the *neo* gene still confer kanamycin resistance in bacteria¹¹ and in plants¹². Plasmid pLB884 (ref. 9) contains the truncated gene *bt884* and encodes an NH₂-terminal fragment of Bt2 up to amino-acid position 610. In *E. coli* it produces a polypeptide of the expected size which is fully toxic towards insect larvae⁹.

Fig. 1 *a*, Structure of the *bt2* gene. The smallest gene fragment encoding an active insect toxin is indicated, and the 3' end positions (codon numbers) of the different truncated genes. *b*, Chimaeric genes derived from *bt2* present in plant expression vectors. The 5' end of the *bt2* coding sequence is fused to the 2' promoter fragment of the TR DNA. pGS1161 contains the intact *bt2* gene. The *bt2* segment in pGS1163 ends at nucleotide position 1,830 of the *bt2* coding sequence. In pGS1151 and pGS1152, the *neo* gene has been fused to 5' fragments of the *bt2* gene at positions 2,173 and 2,050 of the *bt2* coding sequence, respectively. PTR, a 482-base pair (bp) fragment containing the TR DNA 1' and 2' promoters, isolated from pOP443 (ref. 15); 3't7, a 211-bp fragment containing the polyadenylation site of the TL DNA gene 7 (ref. 23); 3'ocs, a 706-bp *PvuII* fragment containing the polyadenylation site of the octopine synthase gene²⁴. Plasmids pGS1160, pGS1161 and pGS1163 were made using the intermediate plasmid pGSH160 which contains a chimaeric *neo* gene¹⁴. For pGS1151 and pGS1152, containing the *bt2:neo* fusion genes pGSH150, a derivative of pGSH160, lacking the *neo* gene and 3' ocs, was used.



Plasmids pLBK860 and pLBK23 (H.H. *et al.*, in preparation) contain fusion genes *bt:neo860* and *bt:neo23* encoding NH₂-terminal fragments of Bt2 up to amino-acid position 683 and 724, respectively. *E. coli* cells harbouring these plasmids are resistant to kanamycin and produce fusion proteins Bt: NPT860 and Bt: NPT23 which have the expected *M_r* of 106K and 110K respectively and which react with anti-Bt2 (Fig. 2A) and anti-NPTII antibodies. At least 50% of the fusion proteins is present in a soluble form in the bacterial cells. The neomycin phosphotransferase activity of the fusion proteins, as determined in an *in situ* assay¹³, is comparable to wild-type NPTII activity. Little or no enzymatic activity was exhibited by polypeptides of lower *M_r* (Fig. 2b). We conclude that the fusion proteins are relatively stable and responsible for the kanamycin-resistant phenotype. Insect assays revealed that, on a molar basis, the fusion proteins exhibit the same toxicity towards *M. sexta* larvae as intact Bt2 protein which has an LD₅₀ value on first instar larvae of 4 ± 2 ng per larva.

Plant expression vectors

The intact and modified *B. thuringiensis* toxin genes were inserted between the T-DNA borders of plant expression vector pGSH160 (for *bt2* and *bt884*) or pGSH150 (ref. 14) (for *bt:neo860* and *bt:neo23*). These plasmids contain the promoter of the 2' gene, a constitutive promoter which directs expression of mannopine synthase in the TR DNA of plasmid pTiA6 (ref. 15). The *Bacillus* genes are followed by a termination signal provided by the 3' end of gene 7 of pTiA6. The resulting plasmids were mobilized into the *Agrobacterium* recipient C58C1 Rif^R pGV2260. The latter contains an octopine Ti plasmid from which the whole T-DNA region has been deleted and replaced by pBR322 (ref. 14). Recombination between pGV2260 and the expression vector through the homologous pBR322 sequences produced Ti plasmids pGS1161, pGS1163, pGS1151 and pGS1152 containing *bt2*, *bt884*, *bt:neo23* and *bt:neo860*, respectively (Fig. 1).

Two approaches were used to increase the probability of obtaining high levels of toxin expression in plants. First, the expression levels directed from the 1' and 2' promoters from the TR DNA were found to be coordinated¹⁵. Consequently we expected that expression of the *neo* gene controlled by the 1' promoter in plants transformed with pGS1161 and pGS1163, would be correlated with transcription of the toxin gene. Second, we anticipated that plant cells transformed with pGS1151 or pGS1152, containing *bt:neo* fusions, would produce fusion

proteins expressing NPTII activity. Selection for high levels of kanamycin resistance would allow us to select directly for transformed clones producing substantial amounts of *B. thuringiensis* protein.

Transformation of tobacco plants

Transgenic tobacco plants were obtained by leaf disk infection^{14,16} of *Nicotiana tabacum* var. Petit Havana SR1 (ref. 17). Shoots resistant to 50, 100 or 200 µg ml⁻¹ kanamycin were selected in all transformation experiments, indicating that the fusion genes indeed confer NPTII activity on transformed plant cells. Individual transformed plants were grown up and subsequently assayed for kanamycin resistance by testing their ability to produce callus from leaf disks on increasing concentrations of kanamycin and the more toxic aminoglycoside antibiotic G418. Most of the transgenic plants expressing an intact NPTII protein produced highly resistant calli, growing on 1,000 µg ml⁻¹ kanamycin and on 100 µg ml⁻¹ G418. In contrast, plants that expressed a Bt: NPTII fusion protein readily fell into different

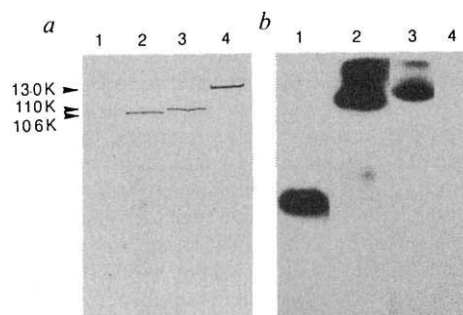


Fig. 2 Antigenic properties and enzymatic activity of Bt: NPTII fusion proteins. *a*, Western blot analysis²⁵ of crude extracts of *E. coli* clones producing intact NPTII enzyme (lane 1), Bt: NPT860 (lane 2) or Bt: NPT23 fusion protein (lane 3). Lane 4, purified Bt2. Blots were incubated with a diluted anti-Bt2 serum and subsequently with alkaline phosphatase-labelled anti-rabbit immunoglobulin. Substrates were 5-bromo-4-chloro-3-indolyl phosphate-toluidine salt and *p*-nitro-blue tetrazolium chloride (Sigma). *b*, Detection of NPTII activity in samples identical to those in *a*, by *in situ* phosphorylation of kanamycin with ³²P-ATP (ref. 13). All samples in *a* and *b* contain 0.25 µg protein per lane.

Fig. 3 Insect toxicity of transgenic tobacco plants expressing *B. thuringiensis* protein. Mortality of *Manduca sexta* larvae after 6 days of feeding on tobacco leaves is shown. The serial number of each plant is given below each column, and in the left-hand panels plants are grouped according to the concentration (in $\mu\text{g ml}^{-1}$) of kanamycin to which they are resistant. Each panel is labelled with the protein expressed (see Fig. 1). Plants were tested 4–6 weeks after transfer to the greenhouse, when they were about 40-cm high and had between six and eight full grown leaves. Two leaf disks of 4 cm in diameter were placed on wet filter paper in Petri dishes and infested with two batches of 10 first instar larvae of *M. sexta*. Leaf disks were replaced daily. Tests were conducted in a growth chamber at 25 °C, 75% relative humidity and under a 16 h light/8 h dark cycle.

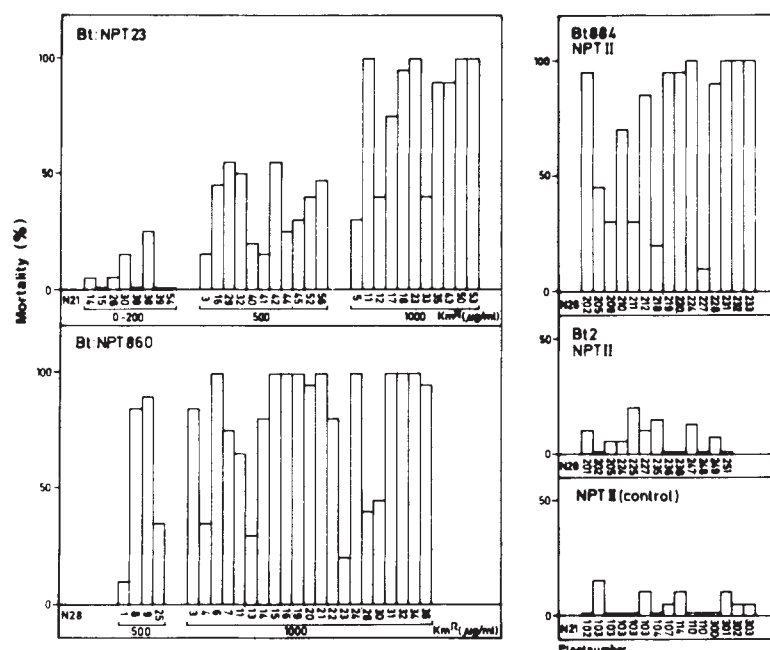


Table 1 Kanamycin and G418 resistance in transgenic tobacco plants

Agrobacterium strain	Expression products of chimaeric genes	No. of plants resistant						G418 ($\mu\text{g ml}^{-1}$)			Total
		<50	50	100	200	500	1000	<10	10	100	
pGS1151	Bt: NPT23	3	1	7	17	14	10	51	1	0	52
pGS1152	Bt: NPT860	0	0	0	0	5	31	35	1	0	36
pGS1161	Bt2, NPTII	0	0	0	0	0	14	0	0	14	14
pGS1163	Bt884, NPTII	0	0	0	0	1	14	1	0	14	15
pGS1160	NPTII	0	0	0	0	2	12	2	0	12	14

Plants were scored according to the highest concentration of antibiotic on which callus could be induced from leaf disks from *in vitro* grown plants²¹.

classes of kanamycin resistance (Table 1). Because the specific enzymatic activity of the fusion proteins is comparable to that of intact NPTII, we presume that these fusion proteins are present in lower amounts in the plant cells than intact NPTII protein.

Insecticidal activity in transgenic plants

Leaves of transgenic plants containing the four types of *B. thuringiensis* gene constructs were fed to *M. sexta* larvae in order to evaluate whether the levels of toxin in the plants would be insecticidal. Mortality rates of *M. sexta* larvae were monitored after 6 days of feeding on leaves of transformed plants (Fig. 3). We found that high toxicity to insects, resulting in 75–100% mortality of the larvae, was observed in about one quarter of the plants that expressed the longer fusion protein Bt: NPT23 and in about two-thirds of those with the shorter Bt: NPT860. Thus, the Bt: NPTII fusions allowed us to select transformants that express levels of the toxin sufficiently high to be insecticidal. Second, insect toxicity caused by the fusion proteins is directly correlated with the level of kanamycin resistance of the transformed plant. In addition, the short fusion generates a larger fraction of transformants expressing high kanamycin resistance and insect toxicity than the longer fusion, suggesting that the shorter *bt: neo860* gene provides higher levels of biologically active protein than the longer *bt: neo23*. Clear insecticidal activity was also detected in most of the 15 plants expressing the truncated *bt884* gene, of which two-thirds induced more than 75% larval death. None of the plants transformed with the full length *bt2* gene produced insect killing activity above levels

obtained in NPTII-expressing control plants. These experiments indicate that for the promoter gene constructs we used, only truncated *bt2* genes give rise to expression levels that are strongly insecticidal in transgenic tobacco.

Protection from insect damage

To test whether expression of modified *bt2* genes in plants results in effective protection against insect damage, selected transgenic plants were grown in the greenhouse and were infested with freshly hatched larvae of *M. sexta*. Plants were kept under conditions that were optimal for survival and growth of the insects. Plants N21-11 and N28-16 were highly protected, because the larvae stopped feeding within 18 hours and all were killed within three days. The damage caused by single larvae was limited to areas of only a few square millimetres (Fig. 4). Other plants, such as N21-53 or N28-6, suffered slightly more damage. On these plants however all the larvae were killed after six days. Control plants such as N21-110, transformed with pGS1160, or untransformed SR1, were severely damaged within 4–6 days and were entirely consumed after 12 days.

Expression of chimaeric genes in plants

Quantitative detection of *B. thuringiensis* toxin in the leaves of transformed plants was performed using a sensitive ELISA, with a mixture of monoclonal antibodies specific for the NH_2 -terminal region of the protein. A correlation was found between the quantity of *Bacillus* protein and insecticidal activity in the transgenic plants. Plants transformed with the truncated *bt2* gene or the fusion constructs contain approximately ten times

more *Bacillus* protein than those transformed with the complete *bt2* sequence (Table 2). Thus, the failure to obtain insect-resistant plants using the intact *bt2* gene is most probably due to inefficient protein synthesis in these transformed plant cells.

Transgenic plants that express the shorter Bt: NPT860 protein are on average more effective in killing insects and express higher levels of toxin than those expressing the longer Bt: NPT23 protein (Table 2). The Bt: NPTII fusion proteins detected in leaf extracts of the transformed plants had the expected size, as determined in Western blots. NPTII enzyme activity was exclusively associated with these fusion proteins as determined by *in situ* NPTII assays. Production of incomplete proteins or degradation in the plant cells was not observed. Thus, transformed tobacco plants produced fully functional fusion proteins of the size of the intact gene product.

The steady-state *Bacillus* messenger RNA levels in the transgenic plants were low and could not be reliably detected in Northern blot analysis. Therefore, they were quantified using ribonuclease protection experiments¹⁸. A probe containing the 5' fragment of the *bt2* coding sequence up to nucleotide position 186 was synthesized using the SP6 transcription system and annealed to total RNA from the leaves of transgenic plants. After ribonuclease digestion, the protected fragments were run on a denaturing polyacrylamide gel. The results showed that RNA levels in the leaves of the transformed plants were correlated with *B. thuringiensis* protein levels (data not shown). The *B. thuringiensis* mRNA in N28-16, the plant that produces the highest level of protein, corresponds to ~0.0001% of the poly(A)⁺ mRNA. The *Bacillus* protein detected in this plant represents 0.02% of the total soluble protein, or 3 µg of this protein per gram fresh leaf tissue. A fivefold lower level, as present in plant N28-34, was sufficient to induce 100% killing in a 6-day insect assay (Table 2). For comparison, the expression levels of the intact *neo* gene also driven by the TR 2' promoter, varied between 10 and 40 µg of NPTII per gram leaf tissue (0.07–0.27% of total protein). This difference in amount of protein is consistent with the lower kanamycin resistance levels of plants transformed with the *bt: neo* fusions compared to those transformed with the intact *neo* gene (Table 1).

Inheritance of the protection

Eighty-five transgenic tobacco plants, transformed with the four types of chimaeric *B. thuringiensis* genes and expressing various levels of active insect toxin, were grown in the greenhouse. All grew normally and were indistinguishable from controls in morphology and vigour of growth. We analysed the inheritance of the kanamycin resistance in eleven plants expressing *bt: neo* genes. Most of these plants contained one (N21-23, N21-50, N28-31, N28-34, N28-21) or two (N21-35, N28-19, N28-24, N28-32) kanamycin resistance loci, as confirmed by Southern blot analysis. Interestingly, some plants that were recognized as producing a large amount of *Bacillus* protein, such as N21-11 and N28-16, generated exclusively kanamycin-resistant F₁ seedlings. DNA analysis showed the presence of at least five copies of the T-DNA in these plants.

F₁ progeny from some transgenic plants were assayed for the expression of insecticidal activity. Insect toxicity was correlated with the kanamycin resistance marker in the F₁ progeny of plants N21-23, N21-50 and N28-34. Insecticidal activity was similar to that observed in the parental plants. Approximately 15 F₁ progeny of N21-11 and N28-16 were analysed. They all induced 100% killing of *M. sexta* larvae in the standard 6-day assay. Toxin levels in the F₁ progeny of N21-11 varied between 20 and 50 ng per mg of total protein, comparable to the 30 ng per mg in the parental plant.

Discussion

Four chimaeric genes containing modified *Bacillus* toxin genes under the control of the 2' promoter of the *Agrobacterium* TR DNA, have been transferred into tobacco plants. All contain

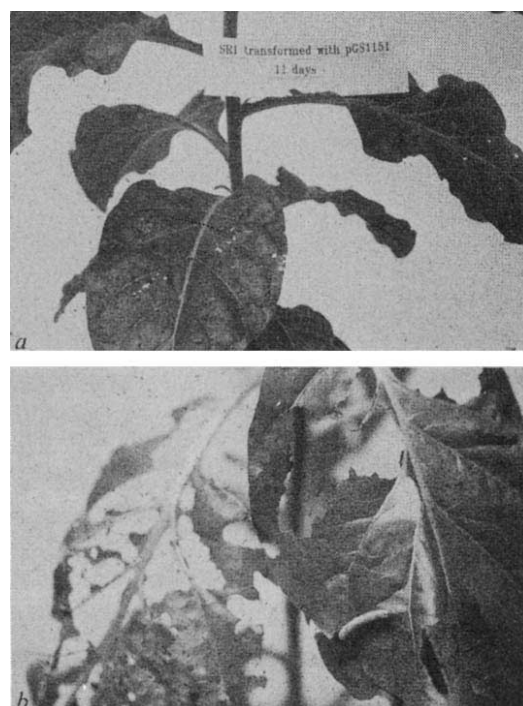


Fig. 4 Protection from insect feeding damage in transgenic plants expressing *B. thuringiensis* protein. Plants 40–50-cm high were infested in the greenhouse with fifteen *M. sexta* larvae per plant. **a**, On N21-11, expressing Bt: NPT23, all larvae died within three days. Leaf damage is very limited. **b**, Damage on a control plant expressing NPTII. Pictures were taken after 11 days.

the toxic core of the Bt2 protein; *bt2* encodes the complete *M*, 130K protoxin, *bt884* is a 5' fragment of *bt2* up to codon 610. Bt: *neo23* and Bt: *neo860* encode fusion proteins which are relatively stable, both in bacteria and plants, and which retain full insect toxicity and NPTII enzyme activity.

Insecticidal levels of toxin were produced when truncated *Bacillus* genes or fusion constructs were expressed in transgenic plants. Mortality rates among *M. sexta* larvae feeding on transgenic leaf material depended on the amount of toxic polypeptide produced. Typically, greenhouse grown plants producing more than 0.004% of their protein as the toxin produced 100% mortality in 6-day feeding assays. Some of the plants we have selected, contain toxin at three to five times this level (N21-11 and N28-16). In greenhouse tests, these plants were well protected from leaf damage caused by insects. Quantification also showed that the toxin expressed in plants has the same specific activity as in a bacterial host.

No significant insecticidal activity could be obtained using the intact *bt2* coding sequence, despite the fact that the same promoter was used to direct its expression. Intact Bt2 protein and RNA amounts in the transgenic plant leaves were 10–50 times lower than those for the truncated *B. thuringiensis* polypeptide or the fusion proteins. Expression levels were not significantly influenced by fusing the *neo* gene to the *bt2* sequence, but rather by the length of the *bt2* fragment. Why the complete *bt2* gene is not expressed at an equally high level in plant cells, is not known. Several parameters, such as differential RNA stability and translation efficiency might be important.

We observed in transgenic plants containing the *bt: neo* fusion constructs a correlation between insecticidal activity and resistance to kanamycin. Three-quarters of all plants resistant to 1,000 µg ml⁻¹ kanamycin induced 75–100% insect mortality. Such fusion genes can be used to select efficiently for transformed plants expressing strong insecticidal activity through direct selection for high kanamycin resistance.

Table 2 Insecticidal activity and *B. thuringiensis* protein content of transformed plants

Toxic protein	Plant	<i>Bacillus</i> protein detected (ng per mg total protein)	Insecticidal effect (% mortality)			Weight reduction in surviving larvae (%)
			Day 3	Day 4	Day 6	
Bt: NPT23	N21-11	33 (132)	90	100	100	—
	N21-35	6.9 (71)	0	40	90	57
	N21-17	2.6	0	15	75	71
	N21-18	5.7	25	50	90	80
	N21-32	2.5	5	40	50	39
	N21-41	4.3	0	0	15	34
	N21-43	4.7	10	40	80	66
Bt: NPT860	N28-16	42 (190)	100	100	100	—
	N28-34	6.9 (42)	70	85	100	—
	N28-6	13	80	90	100	—
	N28-15	10	75	95	100	—
	N28-19	6.2	35	65	100	—
	N28-21	7.0	45	75	100	—
	N28-24	12	60	90	100	—
	N28-31	6.3	55	80	100	—
	N28-32	14	85	90	100	—
Bt2	N21-105	1.3 (5.5)	0	5	15	7
	N21-225	1.2 (2.1)	5	10	20	17
	N21-201	<1	0	5	10	23
	N21-236	1.3	0	0	0	0
	N21-238	<1	0	0	0	0
	N21-249	1.8	0	10	10	10
Bt884	N28-212	30 (125)	100	100	100	—
	N28-220	11 (40)	60	95	100	—
	N28-219	11	55	65	95	85

Values in parentheses refer to amounts detected in greenhouse grown plants. Extracts were prepared from leaves of propagated *in vitro* plants or from leaves of plants grown in a greenhouse that had between six and eight fully expanded leaves. Leaf tissue was ground up and subsequently sonicated (10 s at 50 W) in extraction buffer (Na₂CO₃, 50 mM at pH10; dithiothreitol (DTT), 5 mM; leupeptin, 1 mg ml⁻¹; Triton X-100, 0.05%; EDTA 50 mM; phenylmethylsulphonyl fluoride (PMSF), 0.19 mg ml⁻¹). The extract was cleared by centrifugation and *B. thuringiensis* polypeptides in the supernatant were quantified using an indirect enzyme-linked immunosorbent assay²² (ELISA). Polyvinyl microtitre plates were coated with a goat antibody against *B. thuringiensis* crystal protein. Plant extract dilutions were incubated at 4 °C for 2 h in the coated wells. After rinsing, bound antigen was reacted with a mixture of four distinct monoclonal antibodies against Bt2 and subsequently with an alkaline phosphatase conjugated goat anti-mouse immunoglobulin antibody. The bound enzyme conjugate was detected by adding *p*-nitrophenyl phosphate as a substrate, and relative quantities were determined by measuring absorbance values at 405 nm. The monoclonal antibodies used specify antigenic epitopes located between amino-acid positions 29 and 222 in the NH₂-terminal region of the Bt2 protein (H. Vanderbruggen *et al.*, in preparation). To quantify the *Bacillus* protein levels in transgenic plants, ELISA binding curves of leaf extracts were compared to standard binding curves, obtained by diluting known quantities of the purified homologous protein in control extracts from non-transformed SR1 tobacco plants. The detection limit of the test for purified, solubilized Bt2, was 0.1–1.0 ng ml⁻¹. Toxin levels in plant leaves are expressed as ng toxin per mg of total soluble protein. Insecticidal activity was determined using *Manduca sexta* larvae feeding on leaf material. Mortality after 3, 4, 6 days was determined and weight reduction in the surviving larvae measured after 6 days.

Our experiments illustrate the feasibility of engineering plants that defend themselves against lepidopteran insects which are sensitive to the *B. thuringiensis* berliner insect toxin. However some species, such as *Heliothis* and *Spodoptera* which belong to the *Noctuidae*, an important group of pest insects, are less sensitive to common strains of *B. thuringiensis*, including berliner 1715 (ref. 19). To protect plants fully against these insects, higher levels of expression will be required. This might be achieved using chimaeric *Bacillus* genes containing stronger plant-specific promoters. The 35S promoter of cauliflower mosaic virus²⁰, for example directs a 10–50-fold higher expression than the regular T-DNA promoters in plants. Alternatively, it may be possible

to construct chimaeric toxin genes with higher specific activity against important target insects. Transfer of different chimaeric genes into a variety of crops may provide a new and environmentally safer method of controlling destructive insect pests.

We thank Hilde Van De Wiele, Anne De Sonville, Veronique Gosselé, Jan Dockx, Annemie Van Houtven and Carine Matthijs for their contribution to the experiments. We also thank Leonor Fernandez for her contribution to the initial experiments, Danny Degheele for advice on insect tests, Karin Tenning for preparation of the manuscript, and Karel Spruyt and Stefaan Van Gijsegem for the figures.

Received 3 March; accepted 31 April 1987.

- Dulmage, H. T. *et al.* in *Microbial Control of Pest and Plant Diseases 1970–1980* (ed. Burges, H. D.) 193–222 (Academic, London, 1980).
- Yamamoto, T. & McLaughlin, R. E. *Biochem. biophys. Res. Comm.* **103**, 414–421 (1981).
- Goldberg, L. J. & Margalit, J. *Mosquito News* **37**, 355–358 (1977).
- Krieg, V. A., Huger, A. M., Langenbruch, G. A. & Schnetter, W. Z. *Angew. Ent.* **96**, 500–508 (1983).
- Hernstadt, C., Soares, G. G., Wilcox, E. R. & Edwards, D. L. *Biotechnology* **4**, 305–308 (1986).
- Calabrese, D. M., Nickerson, K. W. & Lane, L. C. *Can. J. Microbiol.* **26**, 1006–1010 (1980).
- Lilley, M., Ruffell, R. N. & Sommerville, H. J. *gen. Microbiol.* **118**, 1–11 (1980).
- Gheysen, G., D'haese, P., Van Montagu, M. & Schell, J. in *Genetic Flux in Plants* (eds Hohn, B. & Dennis, E. S.) 11–47 (Springer, New York, 1985).
- Höfte, H. *et al.* *Eur. J. Biochem.* **161**, 273–280 (1986).
- Herrera-Estrella, L. *et al.* *EMBO J.* **2**, 987–995 (1983).

- Reiss, B., Sprengel, R. & Schaller, H. *EMBO J.* **3**, 3317–3322 (1984).
- Schreier, P., Seftor, E. A., Schell, J. & Bohnert, H. J. *EMBO J.* **4**, 25–32 (1985).
- Reiss, B., Sprengel, R., Will, H. & Schaller, H. *Gene* **30**, 217–223 (1984).
- Deblaere, R. *et al.* *Meth. Enzym.* (in the press).
- Velten, J., Velten, L., Hain, R. & Schell, J. *EMBO J.* **3**, 2723–2730 (1984).
- Horsch, R. *et al.* *Science* **227**, 1229–1231 (1985).
- Maliga, P., Sc. Breznovitis, A. & Marton, L. *Nature new Biol.* **244**, 29–30 (1973).
- Zinn, K. *et al.* *Cell* **34**, 865–879 (1983).
- Kalfon, A. R. & de Barjac, H. *Entomophaga* **30**, 177–186 (1985).
- Odell, J. T., Nagy, J. & Chua, N. M. *Nature* **313**, 810–812 (1985).
- Deblaere, R. *et al.* *Nucleic Acids Res.* **13**, 4777–4788 (1985).
- Clark, M. F. & Adams, A. N. *J. gen. Virol.* **34**, 475–483 (1977).
- Velten, J. & Schell, J. *Nucleic Acids Res.* **13**, 6981–6998 (1985).
- Gielen, J. *et al.* *EMBO J.* **3**, 835–845 (1984).
- Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).

Lipoxygenase metabolites of arachidonic acid as second messengers for presynaptic inhibition of *Aplysia* sensory cells

D. Piomelli[†], A. Volterra[†], N. Dale^{*}, S. A. Siegelbaum^{*†},
E. R. Kandel^{*‡}, J. H. Schwartz^{*§||} & F. Belardetti^{*¶}

^{*} Howard Hughes Medical Institute and Center for Neurobiology and Behavior, Departments of [†] Pharmacology, [‡] Psychiatry, [§] Physiology and ^{||} Neurology, Columbia University, College of Physicians and Surgeons, New York, New York 10032, USA

Biochemical and biophysical studies on Aplysia sensory neurons indicate that inhibitory responses to the molluscan peptide FMRFamide are mediated by lipoxygenase metabolites of arachidonic acid. These compounds are a new class of second messengers in neurons.

IN many of the synaptic actions of neurotransmitters, the binding of transmitter to a cell membrane receptor alters the activity of ion channels indirectly through a second messenger. The number of substances that have been recognized as second messengers in neurons continues to increase since Sutherland's discovery of cyclic AMP in 1957¹ and now includes, in addition to cAMP²⁻⁴, cGMP⁵, Ca²⁺ (ref. 6) and the phospholipid metabolites, diacylglycerol⁷⁻¹⁰ and inositol phosphates⁹⁻¹¹. In addition to these diffusible second messengers, recent evidence indicates that GTP-binding proteins¹² can also modulate ion channels¹³⁻¹⁶. Are there other kinds of second messengers yet to be identified?

A promising mechanism for generating second messenger molecules for neuronal signalling is the receptor-mediated release of arachidonic acid from membrane phospholipids. Release of arachidonic acid has been well characterized in non-neural cells, where free arachidonate is rapidly metabolized to a family of bioactive products (the eicosanoids) that are thought to be both intracellular and intercellular messengers (see ref. 17 for review). Although the metabolism of arachidonic acid in the brain has been studied for more than 20 years¹⁸⁻²³, the complexity of the mammalian central nervous system has made it difficult to assign specific functions to the eicosanoids. We have taken advantage of the simple nervous system of the marine mollusc *Aplysia*, where it is possible to examine identified nerve cells of known behavioural function both in intact ganglia and in dissociated cells in culture. Our results indicate that inhibitory synaptic actions of the neuroactive peptide FMRFamide, which produces presynaptic inhibition of transmitter release in sensory neurons of *Aplysia*, are mediated by lipoxygenase metabolites of arachidonic acid.

Previous studies of *Aplysia* sensory neurons have characterized an excitatory effect of serotonin (5-hydroxytryptamine or 5-HT) that involves a prolonged all-or-none closure of a specific class of K⁺ channels (the S channels) through cAMP-dependent protein phosphorylation²⁴⁻²⁷. Closure of these channels by 5-HT produces slow depolarization, broadens the action potential in the cell body and increases transmitter release from sensory neuron terminals (presynaptic facilitation). Recently, application of FMRFamide (a tetrapeptide denoted by FMRF in single-letter code) to these sensory neurons was found to hyperpolarize the membrane, decrease the duration of the action potential²⁸, increase an outward K⁺ current^{29,30} and inhibit synaptic transmission (presynaptic inhibition)²⁸; actions that are opposite in effect to those of 5-HT. In other molluscan neurons, FMRFamide has also been shown to decrease the inward calcium current^{30,31} and to activate an inward cationic current^{32,33}.

Using single-channel recording, Belardetti *et al.*³⁴ showed that FMRFamide increases the net outward K⁺ current in sensory neurons by increasing the probability that S channels are in the open state (see also ref. 30). FMRFamide can also reverse the closure of S channels produced by 5-HT or cAMP³⁴. Thus, the actions of FMRFamide and 5-HT are antagonistic both at the cellular level and at the level of the single channel.

Several independent lines of evidence suggested that the effects of FMRFamide are produced by a second messenger other than cAMP^{29,30,34}. We have examined the arachidonic acid cascade because its metabolites are formed in *Aplysia* neural tissue in response to application of histamine (ref. 35), a neuromodulator capable of producing presynaptic inhibition in other neurons of *Aplysia*³⁶.

Aplysia eicosanoid pathways

Arachidonate is a prominent component of *Aplysia* neural phospholipids (~10% of total fatty-acid content) and can be metabolized in *Aplysia* neurons through both the 12-lipoxygenase and the 5-lipoxygenase pathways, as well as through the cyclooxygenase pathway³⁵. These pathways are shown in Fig. 1a, with the sites of action of the various drugs used in this study. Two stable lipoxygenase products, the hydroxy acids 12-hydroxyeicosatetraenoic acid (HETE) and 5-HETE, were previously identified in neural tissue from *Aplysia*³⁵. In mammals, these hydroxy acids are formed by enzymatic reduction of the corresponding short-lived hydroperoxides (HPETE)²³. The presence of hydroperoxy acids is only inferred in *Aplysia*, because we have not yet isolated these unstable precursors. Both 12-HPETE and 5-HPETE may be metabolized to products other than the hydroxy acids¹⁷. Finally, the cyclooxygenase pathway leads to the synthesis of various prostaglandins, which in *Aplysia* nervous tissue include PGE₂ and PGF_{2α} (ref. 35).

Arachidonate mimics FMRFamide

To investigate the possible physiological function of arachidonic acid metabolites, we first examined whether exogenously applied arachidonic acid can simulate the actions of FMRFamide on the sensory neurons of the abdominal and pleural ganglia (Fig. 2). When applied extracellularly by pressure ejection from a large pipette³⁷, arachidonic acid (4.5–50 μM) simulated the macroscopic actions of FMRFamide, producing a slow hyperpolarization of the membrane (Fig. 2a) associated with an increase in membrane conductance (not shown), decreasing the duration of the action potential (Fig. 2b), and inhibiting synaptic transmission between the sensory neuron and motor neuron (Fig. 2c). These effects were seen both in intact ganglia (Fig. 2b and c) and in isolated neurons in culture³⁸ (Fig. 2a). Eicosa-11-monoenoic acid, a 20-carbon chain fatty acid with one double

¶ Present address: Department of Pharmacology, University of Texas, Health Science Center, Dallas, Texas 75235, USA.

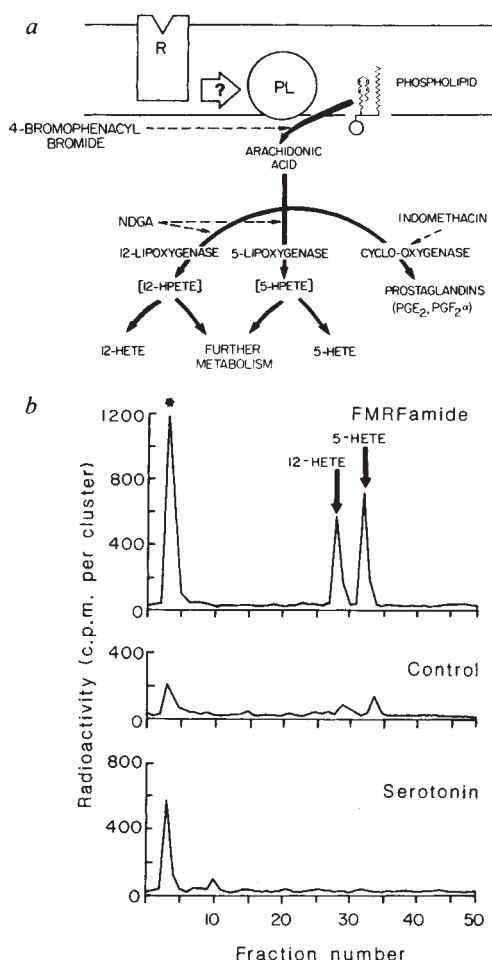


Fig. 1 *a*, Pathways of arachidonic acid metabolism in neural tissue of *Aplysia californica*. In mammals, arachidonic acid release is catalysed by a phospholipase A_2 or by the combined action of a phospholipase C and a diacylglycerol-lipase^{39,40} (PL). Both enzymes can be stimulated in a receptor-mediated (R) fashion and are irreversibly inhibited by 4-bromophenacyl bromide⁴². In *Aplysia*, esterified arachidonate accounts, on average, for 10% of total fatty acids in neural phospholipids³⁵. Free arachidonic acid can be metabolized in *Aplysia* neural tissue through the 12-lipoxygenase, the 5-lipoxygenase and the cyclooxygenase pathways³⁵. Two lipoxygenase products, 12-hydroxyeicosatetraenoic acid (12-HETE) and 5-HETE have been identified³⁵. These stable end-products are thought to be derived from the corresponding unstable hydroperoxy acids (HPETE). The lipoxygenase reaction is selectively inhibited by low concentrations of nordihydroguaiaretic acid (NDGA). *Aplysia* neural tissue also has cyclooxygenase activity and forms several prostaglandins, of which PGE_2 and $PGF_{2\alpha}$ have been identified³⁵. Prostaglandin formation is inhibited by indomethacin. *b*, FMRamide stimulates the release of lipoxygenase products from sensory clusters of *Aplysia*. Representative reversed-phase HPLC fractionation of samples from the incubation of one sensory cluster (≈ 200 cells) with FMRamide (10 μM for 1 min, top trace), without FMRamide (middle) or with 5-HT (10 μM for 1 min, lower trace). Arrows, retention volumes of authentic 12-HETE and 5-HETE standards. The radioactive material associated with the solvent front (*) was consistently increased in the presence of FMRamide (control: 536 ± 136 c.p.m. per 100 cells; FMRamide: $1,701 \pm 377$ c.p.m. per 100 cells, mean $\pm$ s.e.m.; $n = 4$, $P < 0.001$, Student's *t*-test). This component includes polar metabolites of arachidonic acid that are not yet identified.

bond, which constitutes 4.3% of *Aplysia* neural fatty acids³⁵, had little effect on resting potentials, action potentials, or synaptic potentials (Table 1).

The similarity between the physiological effects of FMRamide and arachidonic acid results from their sharing common ionic mechanisms. Voltage-clamp experiments on

Table 1 Comparison of fatty acid metabolites and FMRamide

Resting potential analysis*				
Compound	Concentration (μM)	ΔV_m (mV)	Number responding	<i>n</i>
<i>Cell culture</i>				
FMRamide	2	-6.4 ± 4.9	6	9
Arachidonic acid	4.5–45	-5.8 ± 7.3	4	5
12-HPETE	1.5	-4.6 ± 2.9	3	4
5-HPETE	1.5	0	0	4
12-HETE	1.5	-0.6 ± 0.8	2	5
5-HETE	1.5	0	0	4
<i>Intact ganglia</i>				
Arachidonic acid	10–50	-4.6 ± 5.2	12	15
Eicosa-11-monoenoic acid	50	$+0.2 \pm 0.7$	2	9
Action potential analysis†				
Compound	Concentration (μM)	APD (%)	Recovery (%)	<i>n</i>
FMRamide	5	-18.2 ± 5.7	91.5 ± 7.8	5
Arachidonic acid	50	-17.2 ± 5.4	95.0 ± 3.5	5
Eicosa-11-monoenoic acid	50	$-4.5(-10, +1)$	95.0	2
12-HPETE	60	-27.4 ± 12.4	94.4 ± 9.7	5
5-HPETE	60	-11.8 ± 5.2	96.7 ± 3.9	5
12-HETE	60	-5.7 ± 2.2	98.3 ± 1.6	6
5-HETE	60	-4.3 ± 2.5	98.3 ± 1.9	6
Synaptic potential analysis‡				
Compound	Concentration (μM)	PSP Size (%)	Recovery (%)	<i>n</i>
FMRamide	5–10	-46.8 ± 7.2	85.0 ± 13.9	4
Arachidonic acid	50	-45.0 ± 14.7	84.0 ± 19.0	5
Eicosa-11-monoenoic acid	50	-7.7 ± 1.7	95.7 ± 2.1	3
Single-channel analysis§				
Compound	Concentration (μM)	<i>I</i> (%)	<i>i</i> (%)	<i>n</i>
FMRamide	2–10	$+442.2 \pm 210.4$	-0.06 ± 0.08	5
Arachidonic acid	50	$+269.0 \pm 110.0$	-8.1 ± 5.1	5
Eicosa-11-monoenoic acid	50	+1.2	-9.1	1

* Changes in membrane potential (ΔV_m) as peak response to test compound. Data from sensory neurons in cell culture and intact ganglia as marked. Numbers given are mean $\pm$ s.e.m.; number responding indicates number of cells showing a detectable response (≥ 1 mV). Mean values include responders and non-responders.

† Action potential duration in sensory neurons in intact abdominal ganglia in response to brief suprathreshold current pulses. %APD, maximum per cent decrease in action potential duration from control value in response to test compound; % recovery, per cent return of APD to control value after washing out compound for 5–10 min. Action potential duration measured at time when membrane had repolarized by 80% from peak amplitude.

‡ Postsynaptic potential amplitudes measured in motor neurons from intact pleural ganglia in response to firing the sensory neuron. Maximal % decrease in motor neuron fast EPSP and % recovery are plotted.

§ Single-channel analysis in sensory neurons in cell culture. Mean single S channel current (*I*) measured by integrating 1-min long stretches of data before and after addition of compound to the bath. Maximal % change in *I* and in open channel current amplitudes (*i*) are listed.

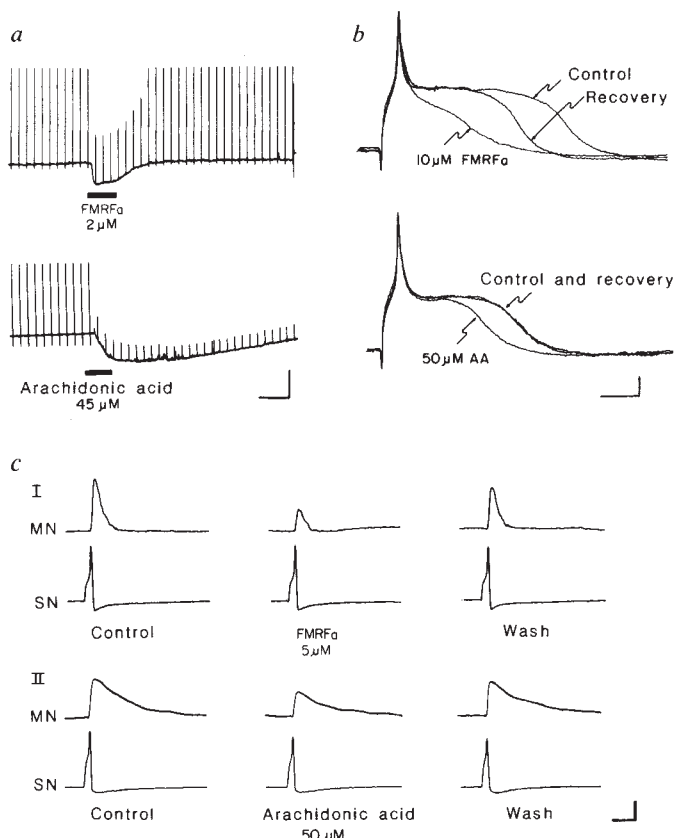
sensory neurons in culture show that arachidonic acid and FMRamide produce similar effects on the membrane current-voltage relation (Fig. 3a). The currents in response to a series of voltage steps, before and shortly after application of FMRamide, are compared in Fig. 3aI. Although the time course of these current traces is complicated by overlapping capacitive-, time- and voltage-dependent ionic currents, FMRamide clearly shifts the current in the outward direction over most of the voltage range.

The FMRamide-sensitive current can be obtained by subtracting the trace recorded in the absence of FMRamide from that recorded in its presence (Fig. 3aII). This FMRamide-sensitive difference current is likely to contain contributions from several different ionic current^{30–34}. At voltages below -30 mV, the difference current displays the characteristic outward rectification typical for the S current^{25,26}. Arachidonic acid produces a similar outward shift in the unsubtracted current record (Fig. 3bI) which yields a similar difference current (Fig. 3bII).

Fig. 2 Arachidonic acid simulates the action of FMRFamide on whole cell potential responses. *a*, Changes in membrane potential and spike firing of sensory neurons in culture in response to FMRFamide (2 μ M, top trace) or arachidonic acid (45 μ M, lower trace). Solid bar, period of application. Intracellular records are from the same cell. Initial resting potential was -41 mV (top) or -35 mV (below). Scale bar: 20 mV vertical, 40 s horizontal. Upward spikes, action potentials elicited once every 10 s by brief depolarizing (constant) current stimuli. The smaller spikes reflect action potential inhibition and are the passive depolarizing responses to the current step. The slower time course of the arachidonate-induced hyperpolarization most probably reflects the time required for diffusion (or transport) across the plasma membrane and metabolism. *b*, FMRFamide and arachidonate (AA) cause a reversible decrease in action potential duration. Action potentials recorded from sensory neuron in an intact abdominal ganglion in presence of 50 mM tetraethyl-ammonium (TEA) (control). In absence of TEA, decrease in action potential duration is also observed with FMRFamide and arachidonic acid; the brief duration of the control action potential makes shortening difficult to measure accurately, however. Resting potential in the top trace was -43 mV before FMRFamide and -48 mV in presence of peptide. In the lower trace, resting potential was -43 mV before arachidonate and -44.5 mV in presence of the fatty acid. Scale bar: 20 mV vertical, 10 ms horizontal. In the intact ganglia, response to arachidonic acid is somewhat slower than in the culture, probably owing to the restricted access and diffusion of the fatty acid. *c*, Reduction in fast excitatory synaptic potential in follower motor neuron (MN traces) in response to firing an action potential in sensory neuron (SN traces). Records I and II are from two different pairs of cells in intact abdominal ganglia. Scale bar: vertical 8 mV for MN, 32 mV for SN traces; horizontal 50 ms.

Methods. Membrane potentials were recorded using conventional intracellular 3 M KCl microelectrodes (10–20 M Ω).

Action potentials were elicited by passing brief depolarizing currents through the recording electrode using a bridge circuit. Two preparations were studied: the gill and siphon mechanoreceptor sensory neurons in intact abdominal ganglia and identified tail mechanoreceptor sensory neurons from the pleural ganglia plated in dissociated cell culture³⁸. Test compounds were applied to sensory neurons by pressure ejection from a wide mouthed pipette³⁷. Concentrations given here were the concentrations of compounds in the pipette. FMRFamide obtained from Peninsula Laboratories. Arachidonic acid obtained from Nu-Check-Prep (Elysian, MN). The artificial sea water (ASW) contained (in mM): 460 NaCl, 10 KCl, 11 CaCl₂, 55 Mg Cl₂, 10 HEPES, pH 7.4. Arachidonic acid in toluene was dried in the dark under a gentle stream of nitrogen immediately before use, ASW added and the mixture sonicated for 15 s. Preparations containing oily droplets, as seen under the microscope, were not used. Arachidonic acid purity was 99%, as measured by thin-layer chromatography (using hexane/ethyl ether/acetic acid/methanol, 85:20:4:2, by volume). Because of the poor solubility of arachidonic acid in water, the free concentration in solution is likely to be lower than the nominal value.



To obtain more direct evidence that FMRFamide and arachidonic acid share a common mode of action, we studied single S-channel currents in cell-attached membrane patches (Fig. 4). S-channels were identified by their characteristic elementary current size, outwardly rectifying current-voltage relation and weak voltage dependence of gating²⁶.

Within 1–2 min of applying either FMRFamide or arachidonic acid in the bath, we observed a large and consistent increase in S-channel opening. The effects on S-channel activity were quantified by measuring the mean (time-averaged) current in the patch (I) and the single-channel current amplitude (i). In the experiment in Fig. 4, FMRFamide increased the mean S-channel current up to sixfold over a range of membrane voltages (Fig. 4b, top) with no significant effect on the single-channel current amplitude (Fig. 4c, top). Arachidonic acid increased mean channel current up to fourfold (Fig. 4b, bottom) with no significant effect on single-channel amplitude (Fig. 4c, bottom). This effect of arachidonic acid was seen in five out of five experiments, whereas eicosa-11-monoenoic acid was ineffective (Table 1).

The mean current through a patch, I , depends on the product $N_i \times i \times p$, where N_i is the number of functional channels in the patch, i the single-channel current amplitude, and p the probability that a channel is open. Previously Belardetti *et al.*³⁴ found that FMRFamide increases I by increasing p , with no effect on N or i . In experiments using a binomial analysis, we were able to confirm that the mode of action of arachidonic acid is similar to that of FMRFamide because it increases p with no marked

effect on N or i . Thus, by all the criteria we have tested, the actions of arachidonic acid and FMRFamide are identical.

Although these experiments show that arachidonic acid simulates the action of FMRFamide, they do not prove that metabolites of endogenous arachidonate mediate the response to the peptide. Arachidonic acid could be acting on the K⁺ channel through an independent parallel mechanism. We have obtained evidence for an obligatory involvement of arachidonic acid using two separate approaches, one pharmacological and the other biochemical.

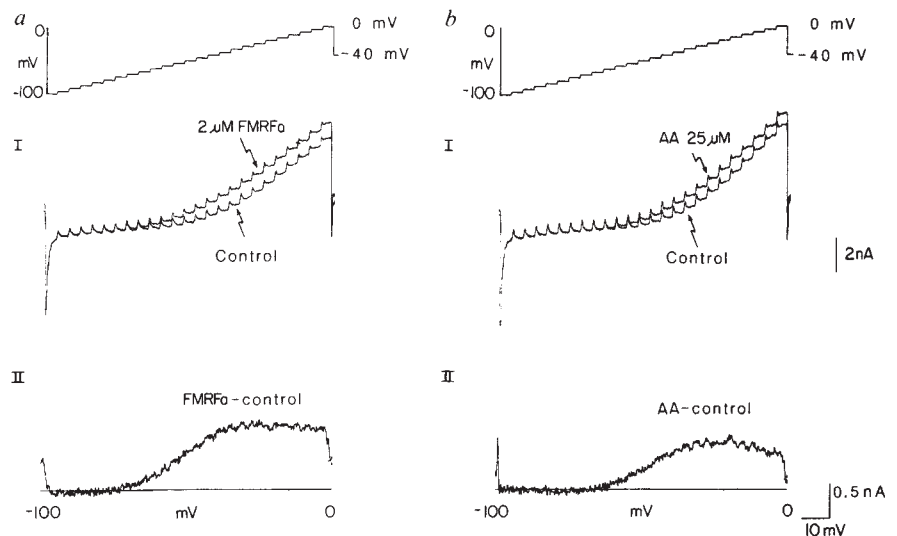
Effects of inhibitors

The first step in the receptor-mediated generation of eicosanoids is the release of free arachidonate from membrane phospholipids, a reaction that in mammals is most commonly catalysed by a phospholipase A₂, or by the combined action of phospholipase C and diacylglycerol lipase^{39,40}. Although selective inhibitors of either phospholipase are not yet available, both enzymes can be effectively blocked by 4-bromophenacyl bromide⁴¹. This inhibitor, at a concentration of 10 μ M, completely suppresses the hyperpolarization produced by FMRFamide in cultured neurons ($n=4$) (Fig. 5b). The depolarizing response to 5-HT is not affected in the same cells, indicating that the drug does not interfere with S-channel function.

The experiments with the phospholipase inhibitor suggest that the release of arachidonic acid is a necessary step in the response

Fig. 3 Comparison of the actions of FMRFamide and arachidonic acid (AA) on the total ionic current in voltage-clamped sensory neurons. *a*, Effect of FMRFamide (2 μ M); *b*, effect of arachidonic acid (25 μ M). Membrane potential held at -40 mV. Once every 30 s, a 500-ms staircase voltage command composed of 20-ms depolarizing steps of 4-mV increments, from -100 mV to 0 mV, was applied (upper traces in *a* and *b*). FMRFamide and arachidonic acid solutions prepared and applied as described in Fig. 2 legend. *a*, Trace I, superimposed records of total membrane current before (lower trace, average of 4 records, control) and at peak action of the peptide (upper trace, average of 2 records, FMRFa). Trace II, the control trace subtracted from the FMRFa trace to yield the FMRFamide-sensitive current. *b*, Current records from the same cell, after prolonged washing with normal sea water. Trace I, superimposed averaged records of total current before (lower trace, average of 5 records, control) and during (upper trace, average of 2 records, AA 25 μ M) application of 25 μ M arachidonic acid. Trace II, arachidonic acid-sensitive current, obtained by subtraction of the records in trace I. Below -30 mV, it seems that the difference current largely reflects the outwardly rectifying S current. Positive to -30 mV contributions from other FMRFamide-sensitive currents (including a decreased calcium and calcium-activated K current) overlap with the S current³⁰⁻³³.

Methods. Sensory neurons in culture were voltage-clamped using a single microelectrode discontinuous voltage-clamp (Axoclamp-2A). Microelectrodes were filled with 2.5 M KCl and had resistances of 10 M Ω . In some experiments, electrodes were coated with a silicone polymer (Sylgard, Dow Corning) to reduce capacitance. Switching frequencies of 5-7 kHz were employed. Current records low pass filtered at 300 Hz using an 8-pole Bessel filter (Frequency Devices), digitized at 1.0 kHz and stored on a hard disk using a DEC 11/73 computer and an INDEC laboratory interface. The computer also applied the voltage clamp command stimuli. Stored current records were averaged and subtracted by computer.



to FMRFamide. To determine which metabolic pathway is responsible for producing the active eicosanoids, we tested the actions of indomethacin (indometacin), which blocks cyclooxygenase⁴² and nordihydroguaiaretic acid (NDGA), which blocks lipoxygenases⁴³. Indomethacin, which we find inhibits formation of prostaglandins in *Aplysia* ganglia with a half-maximal inhibitory concentration of 0.5 μ M without affecting lipoxygenase activity ($IC_{50} > 10 \mu$ M), has no effect on the hyperpolarization induced by FMRFamide (Fig. 5c). In five separate experiments, the response to FMRFamide in the presence of 5 μ M indomethacin was $84 \pm 15\%$ of the control response. In contrast NDGA, which inhibits lipoxygenase activity in *Aplysia* nervous tissue with an IC_{50} of 3 μ M, suppresses the hyperpolarizing response of sensory cells in culture to FMRFamide (Fig. 5d). On average, the hyperpolarization produced by FMRFamide (2 μ M) after treatment with NDGA (5 μ M; $n = 3$) was only $9.3 \pm 7.4\%$ of the response in the absence of the drug. The depolarization produced by 5-HT was not affected. These experiments suggest that a lipoxygenase rather than cyclooxygenase is involved in the intracellular signalling process.

Release of hydroxy acids

To determine whether lipoxygenase metabolites of arachidonic acid were actually released in response to FMRFamide, we studied clusters of sensory neurons dissected from pleural ganglia and labelled by incubation with [³H]arachidonic acid (60 Ci mmol⁻¹). FMRFamide stimulates the formation of radioactive material that migrated on reversed-phase HPLC with authentic 12-HETE and 5-HETE. In four experiments, after a 1-min exposure to FMRFamide (10 μ M), 448 ± 66 c.p.m. per cluster ($\bar{x} \pm s.e.m.$) were recovered in HPLC fractions corresponding to 12-HETE and $1,622 \pm 628$ c.p.m. per cluster in 5-HETE after reversed-phase HPLC purification. A representative chromatogram is shown in Fig. 1b (top trace). In all experiments, the radioactive material at the solvent front also increased in the presence of FMRFamide (control: $1,072 \pm 272$ c.p.m. per cluster; FMRFamide: $3,402 \pm 754$ c.p.m. per cluster, $n = 4$, $P < 0.001$, Student's *t*-test). This component includes more polar

metabolites of arachidonic acid (possibly prostanoids) that are not yet identified. In control incubations (Fig. 1b middle) or incubations with artificial sea water containing 5-HT (10 μ M) (Fig. 1b, bottom), only background radioactivity was found in the 12-HETE and 5-HETE fractions (20 ± 3 c.p.m., $n = 8$).

12-HPETE mimics FMRFamide

Independent support for the involvement of lipoxygenase metabolites comes from experiments where we directly tested the action of exogenously applied metabolites on the sensory neurons. Lipoxygenases convert arachidonate into hydroperoxy acids; in mammals these are either reduced to the corresponding hydroxy acids or transformed into metabolites that are distinctive for each of the lipoxygenase pathways¹⁷. We tested the actions of both 12- and 5- hydroperoxy and hydroxy acids on the resting and action potentials of sensory neurons (Fig. 6). Application of 12-HPETE (1.5 μ M) simulates the slow hyperpolarization of the resting potential and the decrease in excitability seen with FMRFamide (Fig. 6a). 5-HPETE, 5-HETE and 12-HETE have little effect on resting potential (Fig. 6a and Table 1). Both 12-HPETE and 5-HPETE simulate the decrease in action potential duration observed with FMRFamide (Fig. 6b). The average decrease in action potential duration with 12-HPETE, however, was two- to threefold greater than the decrease in response to 5-HPETE. Again, 12-HETE and 5-HETE had little effect on the duration of the action potential (Fig. 6b and Table 1). These results confirm the biochemical and pharmacological experiments that implicate lipoxygenase metabolites in the actions of FMRFamide; they further suggest that the active second messenger derives from a hydroperoxy acid, possibly 12-HPETE, rather than from a hydroxy acid.

Discussion

We find that the receptor-stimulated release of arachidonic acid and its metabolism through a lipoxygenase pathway mediate the inhibitory synaptic response to the tetrapeptide FMRFamide in *Aplysia* sensory neurons. Our results provide strong evidence that lipoxygenase products act as second messengers in nerve

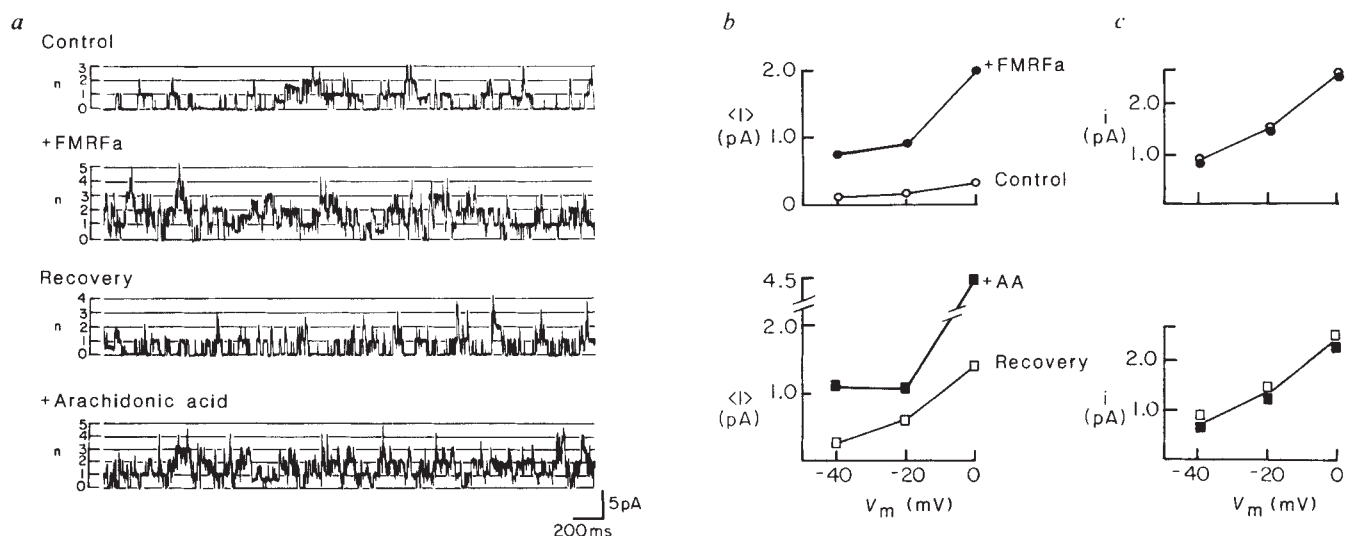
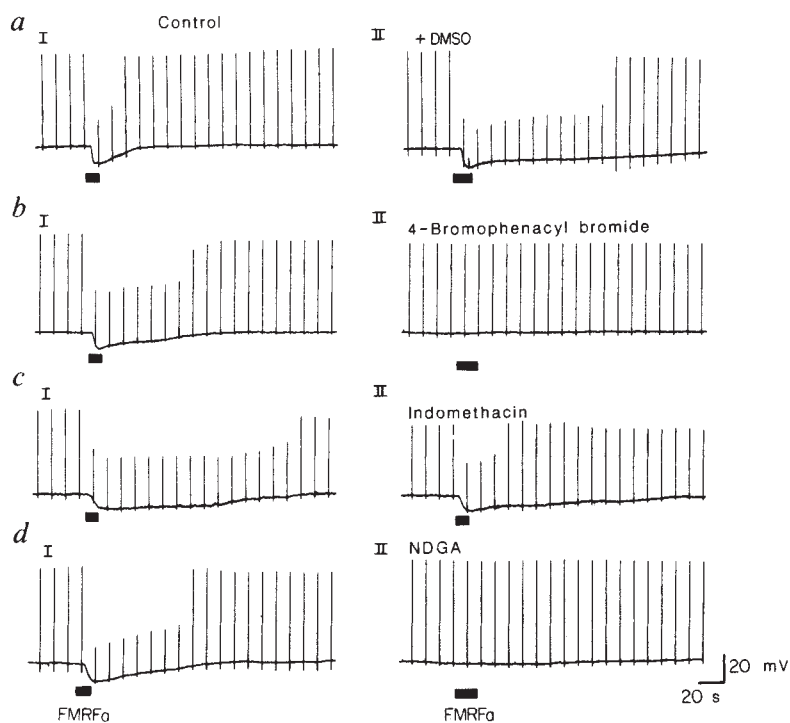


Fig. 4 Comparison of the effects of FMRFamide and arachidonic acid (AA) on single S-channel currents. *a*, Single-channel recordings from a cell-attached patch on a pleural sensory neuron in culture. Control record: before application of any compounds; +FMRFa: 7 min after application of 5 μ M FMRFamide to the bath; recovery: 20 min after washing the peptide out of the bath; +arachidonic acid: 6 min after application of 50 μ M arachidonic acid. Binomial fit to data yields estimate of $n=6$ for the number of active S channels under all conditions. Records obtained at a patch potential depolarized by +80 mV above the resting potential. *b*, Effect of compounds on mean (time-averaged) current through patch. Mean current ($\langle I \rangle$) measured from the time integral of patch current above baseline for 30 s stretches of records at three membrane potentials (V_m). Membrane potential was altered from rest by changing voltage in the patch pipette. Voltages on abscissa are approximate transmembrane potentials of the patch, assuming a resting potential of -60 mV. Current records filtered at 500 Hz. *c*, Effects of compounds on single-channel current amplitude, i . Analysis in *b* and *c* is for the experiment shown in *a*, and symbols are as follows in both *b* and *c*: $\bullet$, +FMRFamide; $\circ$, control; $\blacksquare$, +arachidonic acid; $\square$, recovery.

Fig. 5 Actions of inhibitors of arachidonic acid metabolism on the response to FMRFamide. Chart records of membrane potential under current clamp from sensory cells in culture. Large brief upward deflections are action potentials elicited by depolarizing current pulses. Smaller upward deflections are the passive depolarizations following action potential blockade. Bars, application of 2 μ M FMRFamide onto the cell body. *a-d*, Representative tracings from four different experiments. Left (traces I), control responses to the peptide recorded during continuous superfusion with artificial sea water (resting potential ranged between -38 and -57 mV). Right (traces II), responses to a second application of FMRFamide on the same cells after superfusion of a test drug. *a*, After superfusion with dimethylsulphoxide (DMSO) in artificial sea water (1:10,000 dilution). *b-d*, After superfusion of the sea water/DMSO mixture, containing in addition: in *b*, the phospholipase inhibitor 4-bromophenacyl bromide (10 μ M, Sigma); in *c*, the cyclooxygenase blocker indomethacin (5 μ M, Sigma); in *d*, the lipoxygenase inhibitor nordihydroguaiaretic acid (NDGA, 5 μ M, Biomol). **Methods.** At the beginning of each day, a concentrated stock solution of the blocker in DMSO was prepared, and kept at 0°C in the dark. Portions of this solution were dissolved in artificial sea water to reach the final concentration, vortexed and superfused in the culture dish with a peristaltic pump (at ~ 7 ml min^{-1}). In some of the experiments, indomethacin was dissolved in an ethanol/sea water mixture (1:1,000–1:20,000) or a 100 mM Tris pH 8.4/sea water mixture (1:400–1:1,000).



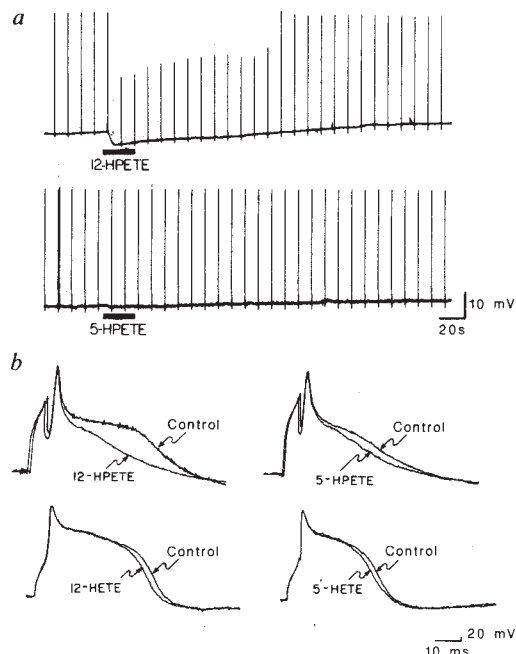
cells and meet the criteria originally proposed for second messenger systems⁴⁴. These are: (1) synthesis—the enzymatic machinery for producing lipoxygenase metabolites is present in neurons; (2) receptor mediation—formation of lipoxygenase metabolites is stimulated by applying the neuropeptide FMRFamide; (3) simulation—arachidonic acid and its lipoxygenase metabolites simulate the actions of FMRFamide; (4) selective blockade—agents that block arachidonic acid release and metabolism effectively inhibit the physiological response to FMRFamide; (5) termination—two mechanisms for terminating the action of the putative second messengers could be reduction

to the inactive hydroxy acids and re-incorporation into membrane phospholipids^{17,35}.

This function for the arachidonic acid cascade is probably not unique to *Aplysia* sensory cells or to FMRFamide, because other experiments show that the presynaptic inhibition produced by stimulating the putative histaminergic L32 cells on the L10-RB connection⁴⁵ operates through the same second messenger mechanism³⁵. It will be of interest to see whether arachidonic acid participates in presynaptic inhibition in vertebrates, such as that produced in dorsal root ganglion cells by noradrenaline⁴⁶ and enkephalin⁴⁷.

Fig. 6 Effect of lipoxygenase metabolites on membrane and action potentials. **a**, Effect of 1.5 μ M 12-HPETE (top trace) and 1.5 μ M 5-HPETE (bottom trace) on resting potentials and stimulated action potentials (brief upward spikes) in pleural sensory neurons in culture. 12-HPETE inhibited spike firing for about 2 min (during this period the smaller spikes are subthreshold). Initial resting potential was -43 mV for top trace and -39 mV for bottom trace. **b**, Effects of metabolites on action potential duration recorded in 50 mM TEA. Action potentials recorded from sensory neuron in intact abdominal ganglia. 12-HPETE and 5-HPETE (60 μ M) were applied to one cell (upper traces); 12-HETE and 5-HETE (60 μ M) were applied to a second cell (lower traces). Action potentials elicited by brief depolarizing stimuli. In the two upper traces the neuron fired a spike after termination of the brief current pulse. Action potentials have been aligned to superimpose the rising phase.

Methods. Lipoxygenase metabolites obtained from Biomol. The hydroperoxy acids were purified by normal phase HPLC using an Alltech silica 4.6 mm $\times$ 25 cm column, eluted with hexane/isopropanol/acetic acid, (98:2:0.01) at 1 ml min $^{-1}$ and stored at -70°C before use. Purity was checked regularly. Immediately before an experiment, a sample of the stock solution was dried under nitrogen, artificial sea water was added and the mixture sonicated.



Eicosanoids differ from other intracellular second messengers in one important way. Unlike other messengers, the eicosanoids are able to leave the cell in which they are generated and act as first messengers on neighbouring cells⁴⁸. For example, a metabolite of arachidonic acid might provide a diffusible extracellular signal from a postsynaptic cell to a presynaptic one. This type of signal could explain the operation of a Hebbian type of synapse of the kind recently postulated to be important for long-term potentiation of pyramidal cells in the CA1 region of the hippocampus, where depolarization of the postsynaptic cell leads to enhanced release from the presynaptic region^{49,50}.

Although our results implicate arachidonic acid metabolites as second messengers for FMRFamide action in the sensory neurons, several questions remain to be answered. First, the inhibitory action of FMRFamide results from a dual ionic mechanism, an increase in outward S current and a decrease in the inward Ca^{2+} current. Our analysis here has concentrated on the increase in S current and we do not know whether the decrease in Ca^{2+} current is also mediated by arachidonic acid metabolites, although the marked decrease in action potential

duration with 12-HPETE and arachidonic acid does suggest a concomitant inhibition of inward calcium current. Second, we have not identified the active metabolite (or metabolites) ultimately responsible for S-channel modulation, although our results indicate that the active molecule is closely related to a hydroperoxy acid. Finally, we do not know either the mechanism coupling transmitter binding to the production of arachidonic acid metabolites or how the active metabolite modulates the channel.

Whatever the answers to these questions, the inhibitory effect of FMRFamide acting through the eicosanoids is functionally antagonistic to the facilitatory action produced by 5-HT through cAMP. Because 5-HT and FMRFamide converge on the same K^+ channel, 5-HT to decrease and the peptide to increase channel activity, our results show that an individual channel can serve as a final common effector for different second messenger systems.

We thank Kathrin Hilten and Louise Katz for preparing the figures and Andrew Krawetz and Harriet Ayers for typing the manuscript. This work was partially supported by a grant from NIH to S.A.S.

Received 10 April; accepted 19 May 1987.

- Rall, W. & Sutherland, R. J. *J. Am. chem. Soc.* **79**, 3608 (1957).
- Siegelbaum, S. A. & Tsien, R. W. *Trends Neurosci.* **6**, 307-313 (1983).
- Kennedy, M. B. A. *Rev. Neurosci.* **6**, 493-525 (1983).
- Nestler, E. J. & Greengard, P. *Nature* **305**, 583-588 (1983).
- Paupardin-Tritsch, D., Hammond, C., Gerschenfeld, H. M., Nairn, A. C. & Greengard, P. *Nature* **323**, 812-814 (1986).
- Rasmussen, H. & Barrett, P. *Physiol. Rev.* **64**, 938-984 (1984).
- De Riemer, S. A., Strong, J. A., Albert, K. A., Greengard, P. & Kaczmarek, L. *Nature* **313**, 313-316 (1985).
- Rane, S. G. & Dunlap, K. *Proc. natn. Acad. Sci. U.S.A.* **83**, 184-188 (1986).
- Higashida, H. & Brown, D. A. *Nature* **323**, 333-335 (1986).
- Hammond, C., Paupardin-Tritsch, D., Nairn, A. C., Greengard, P. & Gerschenfeld, H. M. *Nature* **325**, 809-811 (1987).
- Fisher, S. K. & Agranoff, B. W. *J. Neurochem.* **48**, 909-1017 (1987).
- Gilman, A. G. *Cell* **36**, 577-579 (1984).
- Pfaffinger, P., Martin, J., Hunter, D., Nathanson, N. & Hille, B. *Nature* **317**, 536-538 (1985).
- Breitweiser, G. & Szabo, G. *Nature* **317**, 538-540 (1985).
- Logothetis, D. E., Kurachi, Y., Galper, J., Neer, E. J. & Clapham, D. E. *Nature* **325**, 321-326 (1987).
- Yatani, A., Codina, J., Brown, A. M. & Birnbaumer, L. *Science* **235**, 207-211 (1987).
- Needleman, P., Turk, J., Jakschick, B. A., Morrison, A. R. & Lefkowitz, J. B. A. *Rev. Biochem.* **55**, 69-102 (1986).
- Samuelsson, B. *Biochem. biophys. Acta* **84**, 218-219 (1984).
- Sautebin, L., Spagnuolo, C., Galli, C. & Galli, G. *Prostaglandins* **16**, 985-988 (1978).
- Wolfe, L. S. *J. Neurochem.* **38**, 1-14 (1982).
- Lindgren, J. A., Hökfelt, T., Dahlén, S. E., Patrono, C. & Samuelsson, B. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6212-6216 (1984).
- Shimizu, T., Takusagawa, Y., Izumi, T., Ohishi, N. & Seyama, Y. *J. Neurochem.* **48**, 1541-1546 (1987).
- Jelsema, C. L. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **84** (in the press).
- Kandel, E. R. & Schwartz, J. H. *Science* **218**, 433-443 (1982).

- Klein, M., Camardo, J. S. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5713-5717 (1982).
- Siegelbaum, S. A., Camardo, J. S. & Kandel, E. R. *Nature* **299**, 413-417 (1982).
- Shuster, M. J., Camardo, J. S., Siegelbaum, S. A. & Kandel, E. R. *Nature* **313**, 392-395 (1985).
- Abrams, T. W., Castellucci, V. F., Camardo, J. S., Kandel, E. R. & Lloyd, P. E. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7956-7960 (1984).
- Ocorr, K. A. & Byrne, J. H. *Neurosci. Lett.* **55**, 113-118 (1985).
- Brezina, V., Eckert, R. & Erxleben, C. *J. Physiol., Lond.* **382**, 267-290 (1987).
- Colombaioni, L., Paupardin-Tritsch, D., Vidal, P. P. & Gerschenfeld, H. M. *J. Neurosci.* **5**, 2533-2538 (1985).
- Ruben, P., Johnson, J. W. & Thompson, S. *J. Neurosci.* **6**, 252-259 (1986).
- Cottrell, G. A., Davies, N. W. & Green, K. A. *J. Physiol., Lond.* **356**, 315-333 (1984).
- Belardetti, F., Kandel, E. R. & Siegelbaum, S. A. *Nature* **325**, 153-156 (1987).
- Piomelli, D., Shapiro, E., Feinmark, S. J. & Schwartz, J. H. *J. Neurosci.* (in the press).
- Kretz, R., Shapiro, E. & Kandel, E. R. *J. Neurophysiol.* **55**, 131-146 (1986).
- Choi, D. W. & Fischbach, G. D. *J. Neurophysiol.* **45**, 605-620 (1981).
- Rayport, S. & Schacher, S. J. *J. Neurosci.* **6**, 759-763 (1985).
- Van den Bosch, H. *Biochem. biophys. Acta* **604**, 191-246 (1980).
- Irvine, R. F. *Biochem. J.* **204**, 3-16 (1982).
- Flower, R. J. *Br. med. Bull.* **39**, 260-264 (1983).
- Shen, T. Y. & Winter, C. A. *Adv. Drug. Res.* **56**, 665-675 (1977).
- Egan, R. W. & Gale, P. H. in *Prostaglandins, Leukotrienes and Lipoxins* (ed. Bailey, J. M.) 593-607 (Plenum, New York, 1984).
- Robison, G. A., Butcher, R. W. & Sutherland, E. W. *Cyclic AMP* (Academic, New York, 1971).
- Kretz, R., Shapiro, E. & Kandel, E. R. *J. Neurophysiol.* **55**, 113-130 (1986).
- Dunlap, K. & Fischbach, G. D. *Nature* **276**, 837-838 (1978).
- Mudge, A., Leeman, S. & Fischbach, G. D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 526-530 (1979).
- Hedqvist, P. A. *Rev. Pharmac. Tox.* **17**, 259-279 (1977).
- Wigström, H., Gustafson, B., Huang, Y.-Y. & Abraham, W. D. *Acta physiol. Scand.* **126**, 317-319 (1986).
- Bliss, T. V. P., Douglas, R. M., Errington, M. C. & Lynch, M. A. *J. Physiol., Lond.* **377**, 391-408 (1986).

Circumstellar matter and the nature of the SN1987A progenitor star

R. A. Chevalier* & C. Fransson†

* Department of Astronomy, University of Virginia, PO Box 3813, Charlottesville, Virginia 22903, USA

† Stockholm Observatory, S-133 00 Saltsjöbaden, Sweden

The radio observations of the supernova SN1987A^{1,2} can be interpreted in terms of its interaction with circumstellar matter. The early turn-on of the radio emission implies a relatively low density circumstellar medium, with $\dot{M}/V_w = 8.8 \times 10^{-6} M_\odot \text{ yr}^{-1}$ per (550 km s⁻¹) where $\dot{M}$ is the mass loss rate from the progenitor star and V_w is the wind velocity. The optical properties of the supernova imply that the progenitor star had a smaller radius than that of a typical type II supernova progenitor. The mass loss properties are consistent with this hypothesis. We predict the thermal X-ray luminosity of the supernova and note that it is below the current upper limit³. A bright infrared dust echo is not expected, although a weak echo from an earlier mass loss phase is possible. Weak ultraviolet emission lines from circumstellar gas may be visible. Although the circumstellar density is low, it is possible that the progenitor star did lose a substantial fraction of its mass prior to the supernova explosion.

The bright type II supernovae SN1979c and SN1980k were observed at a range of wavelengths from radio to X-ray wavelengths. Many of the observations can be interpreted in terms of a model involving the interaction of the supernova with a dense circumstellar medium⁴⁻⁶. The radio emission is associated with shocked gas between the expanding supernova and the circumstellar gas. Particle acceleration and magnetic-field enhancement occur in this region, causing synchrotron emission. At early times, free-free absorption by the circumstellar medium gave a low-frequency turnover to the radio spectrum, leading to an estimate of the density of the surrounding medium. Assuming a constant rate of mass loss and a constant stellar wind velocity, Lundqvist and Fransson⁷ find 12×10^{-5} and $3 \times 10^{-5} M_\odot \text{ yr}^{-1}$ per (10 km s⁻¹) for SN1979c and for SN1980k, respectively. SN1980k was observed as an X-ray source in the first few months after maximum light⁸ and the X-ray luminosity was compatible with thermal emission from the circumstellar interaction⁹. Inverse Compton radiation from photospheric photons scattered up in energy by the radio-emitting electrons is also expected^{8,9}. Both supernovae were observed as bright infrared sources at late times, and the properties of the emission were compatible with radiatively heated dust in the circumstellar wind¹⁰. Finally, SN1979c was observed to have emission lines of highly ionized ions¹¹. The ionization was probably caused by radiation from the circumstellar interaction region¹². All of these observed properties are compatible with the presence of the slow, dusty wind expected around a red supergiant star. These are thought to be the progenitors of normal type II supernovae.

Radio observations of SN1987A have been reported for 25.4 and 26.4 February 1987¹, which are days 2.1 and 3.1 after the explosion as determined by the neutrino burst^{13,14}. On day 3.1, the observed fluxes were 90, 130 and 100 mJy at 0.8, 1.4 and 2.3 GHz, respectively, with 10% uncertainties. The implication of the circumstellar interaction model is that the circumstellar gas is just becoming optically thin at these frequencies. The models of ref. 4 imply that t_{20} , the age when an optical depth of unity at 20 cm is reached, is about 2.1 days. For comparison, t_{20} is estimated to be 950 and 190 days for SN1979c and SN1980k respectively⁴. With a maximum supernova velocity of $V_{\text{max}} = 26,000 \text{ km s}^{-1}$ (ref. 15) and a circumstellar temperature $T_c = 1 \times 10^5 \text{ K}$, we obtain $1.7 \times 10^{-7} M_\odot \text{ yr}^{-1}$ per 10 km s⁻¹. This result scales as $V_{\text{max}}^{1.5} T_c^{-0.75}$. The main uncertainty in this estimate is the unknown degree of clumpiness and the assumed wind

temperature. Models for the break-out of the shock wave¹⁶ predict a burst of soft X rays with an effective temperature of 10^5 – 10^6 K , which can heat the gas to a comparable temperature close to the shock⁷. As discussed in ref. 9, it is possible to estimate the thermal and the non-thermal pressures in the shocked region. The energy densities on day 3 are in fact close to those in SN1980k on day 51. The equipartition magnetic field strength is roughly 0.1 gauss. The non-thermal energy density is less than 0.1% of the thermal energy density, so the observed radio luminosity is compatible with the circumstellar interaction model. In this model, the radio luminosity peaks at the time when an optical depth of unity is reached and declines as a power law in time after that. Whereas the model agrees well with the observations during the decline, the rise time may be slower than the exponential predicted, and more compatible with internal free-free absorption². One difference with previous radio supernovae is that Compton cooling of the hot interaction shell may have been important^{9,12}. From the first spectral observations on 25.2 February¹⁷, indicating a photospheric effective temperature of 14,000 K and a luminosity of $\sim 6 \times 10^{41} \text{ erg s}^{-1}$, we find a cooling time of $\sim 2 \times 10^6 \text{ s}$, implying that cooling was only marginal at this epoch. But at phases earlier than this the effective temperature was probably considerably higher, and Compton cooling would be important, giving rise to cold, dense gas immersed within the radio-emitting plasma. This could lead to free-free absorption in the shell, in which case our circumstellar density estimate is an upper limit. The observed radio luminosity, however, indicates that the density is close to our limit.

The deduced circumstellar density for SN1987A is much lower than that found for SN1979c and SN1980k. We believe that this is because the latter two supernovae had red supergiant progenitors and the former did not. The positional coincidence of SN1987A and Sanduleak -69 202 within less than 0.1 arc s¹⁸ strongly argues for this star as the progenitor. This is also consistent with the measured flux in the far ultraviolet one month after the explosion being a factor of ~ 3 lower than that expected, if the star was still present¹⁹. The photometric colours of Sanduleak -69 202 agree well with a B3 Ia supergiant of magnitude $M_v = -6.8$ (ref. 18). From compilations of OB stars in our Galaxy with known mass loss rates^{20,21}, we find that B3 Ia stars have wind velocities of $\sim 550 \text{ km s}^{-1}$ and mass loss rates of $(1.3\text{--}2.8) 10^{-6} M_\odot \text{ yr}^{-1}$. Using the former as a likely value for the velocity we therefore find a mass loss rate of $8.8 \times 10^{-6} M_\odot \text{ yr}^{-1}$ for the progenitor of SN1987A. In view of the sensitivity of this value to T_c and to t_{20} we do not find the discrepancy alarming. Observations also show that OB stars in the Large Magellanic Cloud (LMC) in general, have lower wind velocities than in the Galaxy²², decreasing our determined mass loss rate. From the same compilations we find that the effective temperature of the progenitor should be $\sim 14,500 \text{ K}$, the radius $4.0 \times 10^{12} \text{ cm}$ and the luminosity $1.3 \times 10^5 L_\odot$. Comparing with evolutionary tracks, including mass loss²³, the mass on the main sequence was $\sim 20 M_\odot$. A smaller metallicity may, however, push this estimate lower.

The circumstellar interaction can give rise to X-ray emission, as was observed from SN1980k. As discussed earlier, Compton cooling was ineffective later than day 1, as well as free-free, and the gas was therefore adiabatic. The properties of the thermal emission depend on the density gradient of the supernova gas, we assume that it can be described by a power law, with the power-law index, n , between 7 and 12 (ref. 9). The shocked structure can then be described by a self-similar solution²⁴. The expected temperature, T_x , and luminosity, L_x , of the thermal emission from SN1987A is for $n = 12$, and for $n = 7$

$$T_x = 3.9 \times 10^8 t_{\text{days}}^{-0.4} \text{ K} \quad L_x = 1.1 \times 10^{37} t_{\text{days}}^{-1} A^2 \text{ erg s}^{-1} \quad (n = 7)$$

$$T_x = 1.0 \times 10^8 t_{\text{days}}^{-0.2} \text{ K} \quad L_x = 1.6 \times 10^{38} t_{\text{days}}^{-1} A^2 \text{ erg s}^{-1} \quad (n = 12)$$

where $A = (\dot{M}/V_w)/(8.8 \times 10^{-6} \text{ yr}^{-1} \text{ per } 550 \text{ km s}^{-1})$. The temperature is that of the shocked supernova gas, which is expected

to dominate the emissivity. As the temperature is greater than 2 keV, most of the flux should fall in the detector range of the Ginga Observatory. For the period 25 February to 7 March (day 2 to 13), this observatory set a flux limit to SN1987A of 4 millicrob³, which corresponds to about 6×10^{37} erg s⁻¹. Our model is consistent with the upper limit and predicts that the circumstellar interaction will be difficult to observe in X rays. Inverse Compton X-ray emission is also expected. On day 3, we estimate an inverse Compton luminosity of about 5×10^{35} erg s⁻¹, if the magnetic field is given by its equipartition value. This result, which applies to the range 0.5–5 keV, is uncertain due to uncertainties in the relativistic electron spectrum.

The expansion of the radio-emitting region can be measured using the very-long baseline interferometer technique. This expansion rate depends on the density gradient of the supernova ejecta, on n defined earlier⁹. Using a distance of 52 kpc and a maximum velocity of 26,000 km s⁻¹ for the ejecta at an age of two days, we find that the diameter is given by $0.92 t_{\text{days}}^{0.8}$ marc s for $n = 7$, and $0.79 t_{\text{days}}^{0.9}$ marc s for $n = 12$.

The difference in circumstellar wind properties is likely to result in SN1987A having different infrared properties from the normal type II supernovae discussed above. Because of the low wind density and the high progenitor temperature, dust probably could not form close to the star. But the progenitor star may have been a red supergiant during an earlier phase of evolution. Stellar evolution calculations with mass loss show this possibility²³. The similarity between the H-R diagrams for our Galaxy and for the LMC²⁵, implies that massive stars in the LMC do normally become red supergiants. The faster wind from the more compact star would sweep the dusty red supergiant wind into a shell. If there is a dusty shell of mass M_s a distance R_s from the star, there is infrared emission with a luminosity 5×10^{36} ($E_{\text{rad}}/4 \times 10^{47}$ erg) ($M_s/M_\odot$) (R_s/pc)⁻³ erg s⁻¹ for a time of $6.5 (R_s/\text{pc})$ yr with effective temperature $150 (L_{\text{SN}}/10^{41} \text{ erg s}^{-1})^{0.2} (R_s/\text{pc})^{-0.4}$ K, where E_{rad} is the radiated energy of the supernova and L_{SN} is the supernova luminosity. The dust parameters for the calculation were taken from ref. 10 and it is assumed that the thickness of the dust shell is greater than that of the radiation pulse. For a 50-day light pulse, this implies a thickness greater than 0.04 pc. Because of light travel-time effects the infrared emission would appear as a ring of emission with radius of $(c t(2 R_s - ct))^{0.5}$, where c is the speed of light.

The different properties in the ultraviolet of SN1987A from other type IIs may be explained as a consequence of the low circumstellar density²⁶. For a high density, the ionizing flux from the shock both contributes directly to the ultraviolet continuum and also keeps the abundance of ions, like Fe II, at a low level. As most of the line blocking in the ultraviolet is caused by these ions²⁷, the lower ionization for the parameters of SN1987A can explain the large ultraviolet deficit seen.

Finally, we consider the possible observability of lines from circumstellar medium at late phases. The fluxes of these are sensitive to the temperature and ionization of the gas, and may thus be used as a diagnostic of the very early extreme ultraviolet burst during the first day. Due to the lower background flux the most promising lines are the ultraviolet resonance transitions. A characteristic feature of the lines coming from the circumstellar gas would be the width of the lines, corresponding to the wind velocity, ~ 550 km s⁻¹. For effective temperatures of the extreme ultraviolet burst in excess of $\sim 10^5$ K most ions will, however, be in their He-like stages in the wind region. If the progenitor gave rise to a wind blown bubble, with radius ~ 10 pc, calculations show that the dense shell of swept up gas may possibly be observable in C IV, N V and O VI. As discussed in ref. 26, there is no evidence for absorption lines from the circumstellar region. A wind blown shell would have a velocity of only ~ 10 – 20 km s⁻¹, and therefore difficult to separate from other contributions.

The circumstellar density around SN1987A is small compared to that around normal type II supernovae. The main reason for this is the larger wind velocity of the B3 supergiant progenitor. With a wind velocity of 550 km s⁻¹, the determined mass loss rate implies that several solar masses of matter would be lost over a period of 10^6 yr. Thus it is possible that a substantial fraction of the initial envelope mass was lost. This could be the reason for the small radius of the progenitor star.

We thank P. Lundqvist for discussions on the ionization of the wind. R.A.C.'s research is supported by NASA and the NSF.

Received 22 April; accepted 1 June 1987.

- Bunton, J. D., Turtle, A. J. & Jauncey, D. L. *IAU Circ.* 4321 (1987).
- Turtle, A. J. *et al. Nature* **327**, 38–40 (1987).
- Makino, F. *IAU Circ.* 4336 (1987).
- Chevalier, R. A. *Ann. N.Y. Acad. Sci.* **422**, 215–232 (1984).
- Weiler, K. W., Sramek, R. A., Panagia, N., van der Hulst, J. M. & Salvati, M. *Astrophys. J.* **301**, 790–813 (1986).
- Fransson, C. in *Radiation Hydrodynamics in Stars and Compact Objects* (eds Mihalas, D. & Winkler, K. H. A.) 141–164 (Springer, Berlin, 1986).
- Lundqvist, P. & Fransson, C. *Astr. Astrophys.* (submitted).
- Canizares, C. R., Kriss, G. A. & Feigelson, E. D. *Astrophys. J.* **253**, L17–L21 (1982).
- Chevalier, R. A. *Astrophys. J.* **259**, 302–310 (1982).
- Dwek, E. *Astrophys. J.* **274**, 175–183 (1983).
- Fransson, C. *et al. Astr. Astrophys.* **132**, 1–10 (1984).
- Fransson, C. *Astr. Astrophys.* **133**, 264–284 (1984).
- Koshiha, M. *IAU Circ.* No. 4338 (1987).
- Svoboda, R. *IAU Circ.* No. 4340 (1987).
- Gry, C., Cassatella, A., Wamsteker, W. & Sanz, L. *IAU Circ.* No. 4324 (1987).
- Klein, R. I. & Chevalier, R. A. *Astrophys. J.* **223**, L109–L112 (1978).
- Cassatella, A. *et al. Astr. Astrophys.* **177**, L29–L32 (1987).
- West, R. M., Lauberts, A., Jørgensen, H. E. & Schuster, H.-E. *Astr. Astrophys.* **177**, L1–L3 (1987).
- Panagia, N. *et al. Astr. Astrophys.* **177**, L25–L28 (1987).
- Lamers, H. J. G. L. M. *Astrophys. J.* **245**, 593–608 (1981).
- Underhill, A. & Doazan, V. *B stars With and Without Emission Lines* NASA SP-456 (NASA, Washington DC, 1982).
- Garmany, C. D. & Conti, P. S. *Astrophys. J.* **293**, 407–413 (1985).
- Chiosi, C. & Maeder, A. *Rev. Astr. Astrophys.* **24**, 329–375 (1986).
- Chevalier, R. A. *Astrophys. J.* **258**, 790–797 (1982).
- Humphreys, R. M. & Davidson, K. *Astrophys. J.* **232**, 409–420 (1979).
- Fransson, C., Grewing, M., Cassatella, A., Panagia, N. & Wamsteker, W. *Astr. Astrophys.* **177**, L33–L36 (1987).
- Branch, D. & Venaatkrishna, K. L. *Astrophys. J.* **306**, L21–L23 (1986).

Experimental evidence of the inverse Smith–Purcell effect

K. Mizuno, J. Pae, T. Nozokido & K. Furuya

Research Institute of Electrical Communication, Tohoku University, Sendai 980, Japan

Interest in the development of laser-driven linacs has been stimulated recently by the advent of the high-power lasers. The use of a laser to accelerate charged particles was first proposed by K. Shimoda in 1962¹. He noticed that high values of acceleration per metre (acceleration gradient) could be obtained with an intense electric field in the output of a high-power laser. The inverse Smith–Purcell (ISP) effect proposed by us² is a candidate for laser-driven linacs with a metallic grating as an interaction circuit. In 1953 Smith and Purcell demonstrated that light is emitted when a high-voltage electron beam moves parallel and close to a metallic optical diffraction grating in a direction perpendicular to the grating rulings³. The dispersion relationship is a synchronous condition between the electrons and the wave on the grating. Therefore, the inverse Smith–Purcell effect (ISP), or the extended interaction between electrons and an incident light wave should occur when the same relation is satisfied between the electrons and the incident wave (Fig. 1)². Lawson pointed out⁴ that this effect would fail to accelerate relativistic particles. This difficulty was solved by Palmer⁵, who showed that acceleration is possible either if the particles travel skew to the grating lines, or if the radiation is falling at a skew angle onto the grating. Several authors have described possibilities for producing a grating linac (ISP effect) with accelerating field of a few GeV m⁻¹ in the infrared or optical wave region⁶. Here we report the first observational evidence for

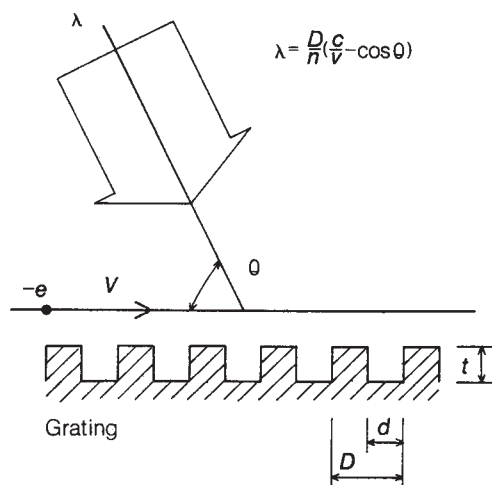


Fig. 1 Schematic of the inverse Smith-Purcell effect.

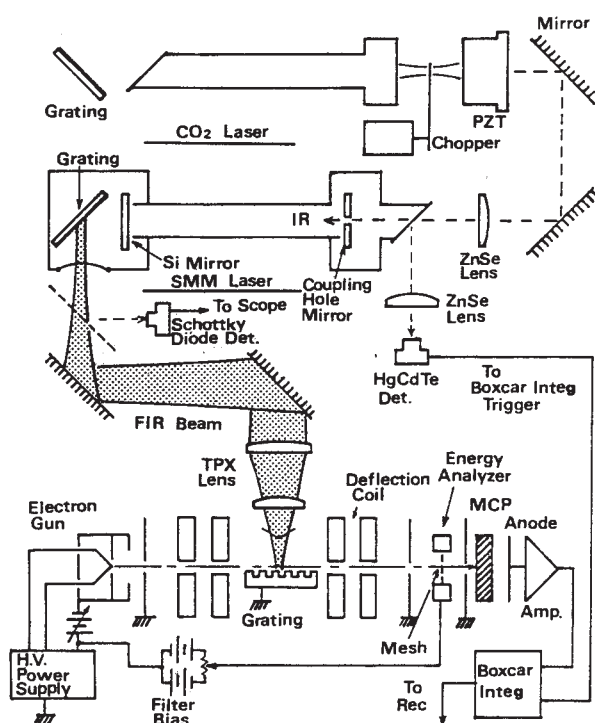


Fig. 2 Experimental set-up to demonstrate the inverse Smith-Purcell effect by using a submillimetre wave laser.

this effect using a submillimetre-wave laser as a driving source. The experimental results give good agreement with theoretical predictions.

We have observed the ISP effect for the first time. One of the features of our experiment is that we used a submillimetre-wave laser as the driving source. Theoretical analysis of the field distribution just in front of the grating shows that a region within 0.1λ (λ , wavelength) of the grating surface may be used for electron interaction⁷. From this point of view, submillimetre rather than infrared or optical wavelengths should be more suitable for the first experiment to demonstrate the existence of the ISP effect because we can use a wider region for the interaction when we use submillimetre wavelength. Figure 2 shows our experimental setup. The driving laser is a submillimetre wave CH_3F laser ($\lambda = 496 \mu\text{m}$) which is excited by an EQ-switched CO_2 laser newly developed for this experiment. The peak power of the submillimetre-wave laser is 2.5 W, the dur-

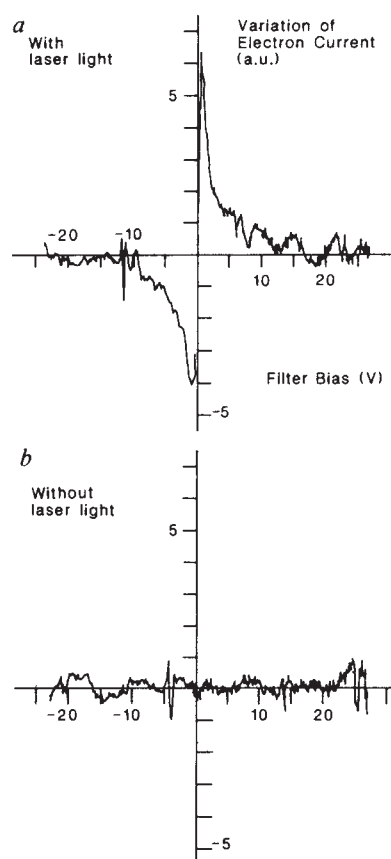


Fig. 3 Experimental evidence for the inverse Smith-Purcell effect. Energy spread of the electron beam is shown as the difference from that of no laser illumination.

ation is 100 ns, and the repetition rate is 500 Hz. The electron energy analyser is of the retarding field type whose resolving power is 0.8 eV full width at half maximum (FWHM) at an electron energy of 80 keV. The groove-dimensions of the grating were designed to give maximum field intensity just in front of the grating. The grating was fabricated on a milling machine. Tolerances of $\pm 2 \mu\text{m}$ on the groove dimensions were maintained using an optical positioning scale and a He-Ne laser.

Figure 3 shows the experimental evidence for the ISP effect. The initial accelerating voltage of the electron beam is 80 kV. The abscissa is the filter bias voltage of the energy analyser, which corresponds to the electron energy spread. The ordinate

is the output of the boxcar averager (shown in Fig. 2), and shows the difference in the electron current with, Fig. 3a, and without, Fig. 3b, laser illumination (the laser pulse is 100 ns in width, and the electron beam is d.c.).

It can be seen that a 2.5-W laser produces an increase in the FWHM of approximately 5 eV. This agrees with the theoretically predicted value within a factor of 3.

A more precise measurement of this effect is now in progress.

Received 9 March; accepted 8 April 1987.

1. Shimoda, K. *Appl. Opt.* **1**, 33-35 (1962).
2. Mizuno, K., Ono, S. & Shimoe, O. *Nature* **253**, 184-185 (1975).
3. Smith, S. J. & Purcell, F. M. *Phys. Rev.* **92**, 1069 (1953).
4. Lawson, J. D. *Rutherford Lab. Rep. Laser Accelerators* RL-75-043 (1975).
5. Palmer, R. B. *Particle Accel.* **11**, 81-90 (1980).
6. Cannell, P. J. (ed.) *AIP Conf. Proc. no. 91* (AIP, New York, 1982).
7. Mizuno, K. & Ono, S. in *Infrared and Millimeter Waves* Vol. 1 (ed. Button, K. J.) (Academic, New York, 1979).

Radio-frequency SQUID operation using a ceramic high-temperature superconductor

M. S. Colclough*, C. E. Gough*, M. Keene*,
C. M. Muirhead*, N. Thomas*, J. S. Abell†
& S. Sutton†

* Department of Physics, The University of Birmingham,
PO Box 363, Birmingham B15 2TT, UK

† Department of Metallurgy, The University of Birmingham,
PO Box 363, Birmingham B15 2TT, UK

The recent discovery¹ of ceramic superconductors with high transition temperatures opens the way for the development of superconducting devices operating at liquid-nitrogen temperature (77 K). Of particular interest is the possibility of constructing a radio-frequency superconducting quantum interference device (r.f. SQUID), for use as an extremely sensitive magnetometer². The operation of the r.f. SQUID is also of interest as a means of investigating superconductivity in the new ceramic materials. Here we report the construction of an r.f. SQUID based on the multi-phase high-temperature superconducting ceramic $Y_{1.2}Ba_{0.8}CuO_4$ (ref. 1). We have used this as a magnetometer operating at 4.2 K and have observed unusual SQUID behaviour which may be due to the presence of many weakly coupled superconducting paths within the material^{1,3}. We have also observed SQUID behaviour at temperatures as high as 45 K.

Conventional r.f. SQUID magnetometers (see ref. 2 for a detailed description of their operation) use a superconducting ring intersected by a 'weak link', where magnetic flux can pass relatively easily in or out of the ring. The ring is magnetically coupled to a tuned LC circuit driven by a constant-current radio-frequency source, and the amplitude of the r.f. voltage across the circuit is detected as a function of the magnetic flux applied to the superconducting ring. For certain levels of r.f.-current bias, the r.f. voltage changes with applied flux owing to a change in the hysteretic losses in the ring. In practice, as the applied flux is varied a triangular waveform is observed on an oscilloscope, with a periodicity equal to the flux quantum $\phi_0 = h/2e$ (where h is Planck's constant and e is the charge of the electron). By 'locking on' to a minimum of the triangular waveform, standard SQUID electronics can maintain a constant flux inside the superconducting loop. This is achieved by feeding a measured d.c. bias current into the r.f. coil sufficient to balance out the effect of any externally applied magnetic field. Such a system then operates as a very sensitive magnetometer, with a typical sensitivity of $10^{-3} \phi_0 \text{ Hz}^{-1/2}$ sufficient even to detect the tiny magnetic fields generated by electrical currents in the brain⁴.

In recent measurements⁵ we have verified that the quantum of magnetic flux, ϕ_0 , in the ceramic superconductor

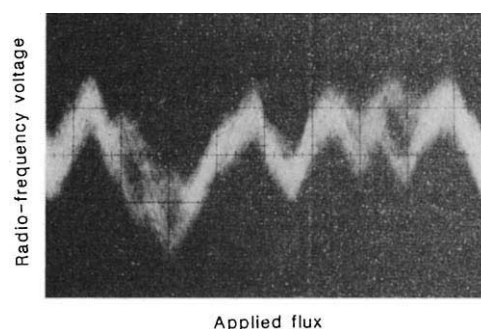


Fig. 1 Typical example of the variation of the r.f. (20 MHz) voltage across the LC circuit with applied magnetic flux, with the specimen at 4.2 K.

$Y_{1.2}Ba_{0.8}CuO_4$ is equal to $h/2e$, demonstrating that high-temperature superconductivity in this material involves the pairing of electrons. Our measurements also showed that a ring prepared with the original multi-phase composition reported by Wu *et al.*¹ behaved as a truly superconducting loop, with inherent weak links within the bulk of the material itself. This suggested that r.f. SQUID operation might be possible simply by substituting a ring of the ceramic material for the superconducting ring and weak link of a conventional SQUID magnetometer², thus eliminating the need for a separate point-contact or fabricated weak-link junction.

An r.f. SQUID magnetometer was therefore devised, which incorporated a ceramic ring of similar dimensions, chemical and physical composition to that used in our earlier experiment⁵. The SQUID was operated at 20 MHz, with the r.f. coil (with outer diameter equal to the inner diameter of the ring) fitted closely against the inner surface of the ring to optimize r.f. coupling, and the flux through the ring was varied sinusoidally at 1 kHz. When operated at 4.2 K in a well-shielded magnetic environment provided by superconducting lead and mu-metal shielding, r.f. SQUID operation was readily observed. But the amplitude of the voltage across the r.f. coil varied with flux threading the ring in a more complicated way than the conventional triangular waveform, and did not exhibit a periodicity corresponding to a single flux quantum threading the ring. Furthermore, it was soon discovered that it was not even necessary to use a ring, and that very similar behaviour could be observed with the r.f. coil wound directly on a solid rod of the material.

The measurements reported here were made with a coil of 25 turns wound on a thin rod of the material of length 8 mm and diameter 1.5 mm. We emphasize that this mode of operation would not be possible for a conventional bulk superconductor, which would simply behave as a perfect diamagnet excluding all flux. The observation of r.f. SQUID behaviour shows that flux can indeed penetrate the multi-phase material to a significant extent, even at high frequencies. To calibrate the magnetic-field sensitivity, the rod and r.f.-coil were placed concentrically inside a long solenoid, which was used to apply an external magnetic field at 1 kHz to the sample. Figure 1 shows a typical example of the variation in amplitude of the r.f. voltage as a function of applied magnetic field. (The field was corrected for screening currents flowing in the superconducting shields.) The change of flux across the sample area, assuming uniform flux penetration, is shown in units of ϕ_0 .

It is possible to interpret the waveform as the superposition of a number of triangular waveforms having a range of periodicities. This is consistent with the existence of a number of superconducting loops of cross-sectional area A , with typical values of about one-twentieth the cross-section of the rod. Any

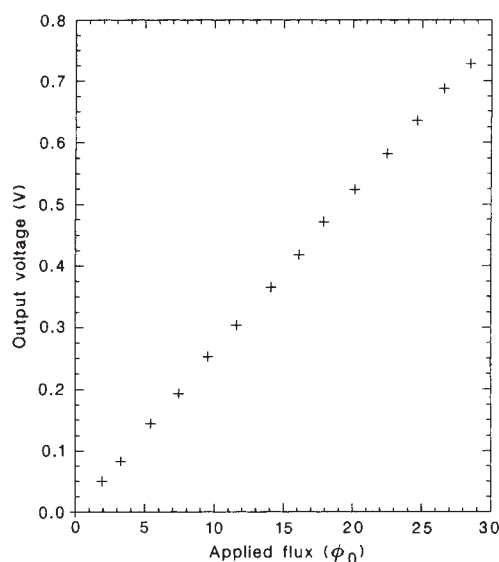


Fig. 2 Magnetometer behaviour at 4.2 K, showing output voltage as a function of the flux applied to the full sample cross-section.

particular loop would then produce a triangular waveform with a periodicity of ϕ_0/A in external field. The sample presumably contains loops of many different areas, and it is therefore surprising that the resulting waveform is not completely random. The roughly periodic waveform may imply that there is a dominant loop area. We also found that the general shape of the waveform was only weakly dependent on the r.f.-current level, in contrast to the discrete levels required for the operation of a conventional r.f. SQUID². The waveform could, however, be changed substantially by thermal cycling, mechanical vibration, or by the application of magnetic fields to the sample. We have not yet developed a detailed explanation for the novel r.f. SQUID behaviour which we have observed, and there are a number of factors, such as the distribution of loop areas and critical currents and the effects of thermal noise and magnetic interactions between the loops, that need to be taken into consideration.

The sensitivity of the waveform to changes in magnetic field enables us to operate the system as a locked-on r.f. SQUID, with the 1-kHz modulation and feedback signal applied to the r.f. coil. But instead of locking on to a minimum of a triangular waveform, one simply locks on to any well-defined minimum in the more complicated waveform. The resulting feedback signal is then directly proportional to any changes in the external field, which in these measurements was applied using the long solenoid. Figure 2 illustrates the sensitivity and linearity of our system over an extended range of field intensity, with the device operating at 4.2 K. The system was stable enough to operate over long periods (>30 min), and had a noise level of better than 0.1 ϕ_0 with a system time constant of 1 ms. The dynamic range is eventually limited by changes in the waveform.

Similar waveforms were observed as the sample was warmed to much higher temperatures, although the more rapidly varying components, associated with loops of larger area, tended to disappear. At 45 K only slowly varying components, corresponding to loops of only one-thousandth the sample cross-sectional area, remained. This might be expected on statistical grounds, as any reduction in the strength of weak links with increasing temperature would result in preferential suppression of those loops containing a large number of weak links relative to the small loops having a smaller number. In addition, thermally generated flux noise, with intensity $(\Delta\phi)^2 \approx LkT$ (where k is the Boltzmann constant and T is the absolute temperature), tends to suppress contributions from the larger loops because of their larger inductance L . High-temperature operation is then

only possible because of the number of small loops within the sample, and the sensitivity of the SQUID to the external magnetic field is, therefore, reduced.

We are currently extending these measurements to other SQUID geometries and different samples in order to optimize the SQUID performance and to clarify the mechanisms for dissipation in these materials.

Received 26 May; accepted 10 June 1987.

1. Wu, M. K. *et al. Phys. Rev. Lett.* **58**, 908–910 (1987).
2. van Duzer, T. & Turner, C. W. *Principles of Superconducting Devices and Circuits* (Elsevier, New York, 1981).
3. Müller, K. A., Takashige, M. & Bednorz, J. G. *Phys. Rev. Lett.* **58**, 1143–1145 (1987).
4. Williamson, S. J. & Kaufman, C. J. *J. Magnetism magn. Mater.* **22**, 129 (1981).
5. Gough, C. E. *et al. Nature* **326**, 855 (1987).

Bubble clouds and temperature anomalies in the upper ocean

S. A. Thorpe & A. J. Hall*

Department of Oceanography, The University, Southampton SO9 5NH,

* Institute of Oceanographic Sciences, Wormley, Godalming GU8 5UB, Surrey,

The processes of mixing in the upper few metres of the ocean are poorly known and largely unquantified¹ but are a vital element in ocean-atmosphere interaction, playing a key part in the transfer and dispersion of heat, gases and nutrients. Here we report results from a towed instrument used to measure simultaneously water temperature and the sound scattered from bubble clouds produced by breaking waves. Dominant features in the records are correlated, intense sound scattering (or bubble clouds) being associated with warmer water and large positive temperature gradients with the onset of increased scattering. The observations provide a more complete picture of the structure and nature of bubble clouds and establish their association with dynamical processes of turbulent diffusion in the upper ocean boundary layer.

An array of 11 thermistors was used to measure the horizontal temperature structure in the upper ocean boundary layer. The array² is mounted on the leading edge of a streamlined spar which is suspended from a catamaran towed from a ship ahead of its wake (Fig. 1) to avoid the effects of ship-induced waves, and provides measurement of temperature at depths of 0.9 to 8.6 m with a resolution of 0.25 mK. The array is sampled at 4 Hz so that, at normal towing speeds of 2–2.5 ms^{-1} when the spar has a mean tilt of $\sim 8^\circ$ from the vertical (it is equipped with tilt and pressure sensors), the effective horizontal sampling interval is 0.5–0.6 m. The spar also carries a forward-pointing 1 MHz sonar transducer with a 5° beam angle at a mean depth of 3.4 m. The pulse repetition frequency is 4 Hz with a pulse length of 0.1 ms. Data proportional to the intensity of sound scattered from targets at nominal ranges of 3.38 m and 9.23 m ahead of the spar are recorded, together with the temperature, tilt and pressure values. (The range is 'nominal', depending on a constant time delay between the emitted and received sound pulses of 4.5 and 12.3 ms, and the speed of sound.) The acoustic returns are also displayed on a sonograph. Wave periods and wind speeds were recorded separately; the pressure sensor also provided a reliable record of wave encounter period through its response to dynamic pressure.

The equipment was deployed four times and towed for a total of 163 h in wind speeds of 0 to 18 ms^{-1} east of Iceland in June 1986. For operational reason unrelated to the instrument, the majority of tows were made into, or nearly into, the wind and therefore reliable data are not available in cross-wind directions. Air temperatures were generally 1–2 K greater than the water. Wind waves with periods of about 3–4 s appeared to dominate over swell.

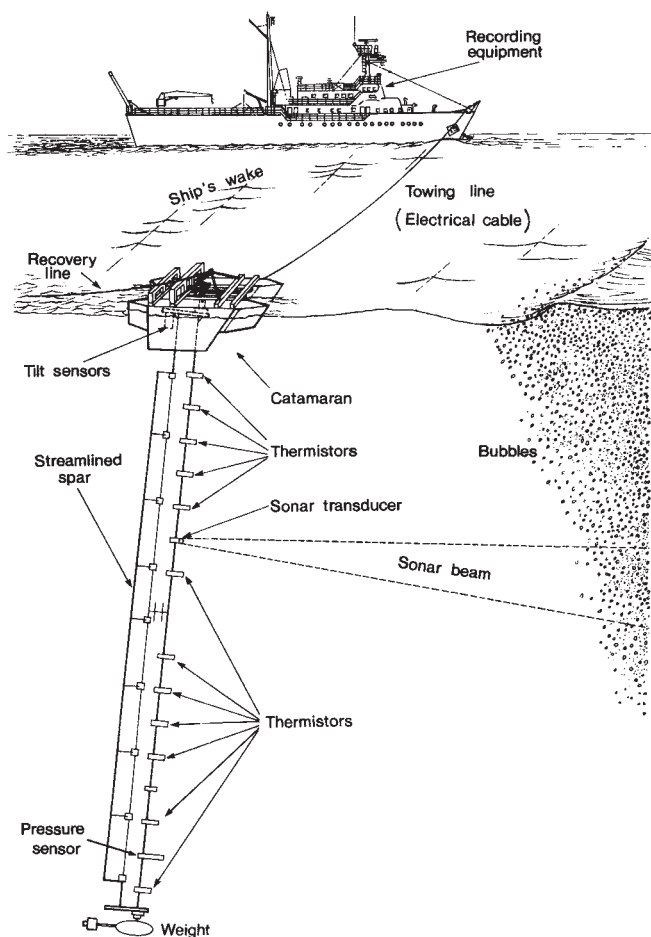
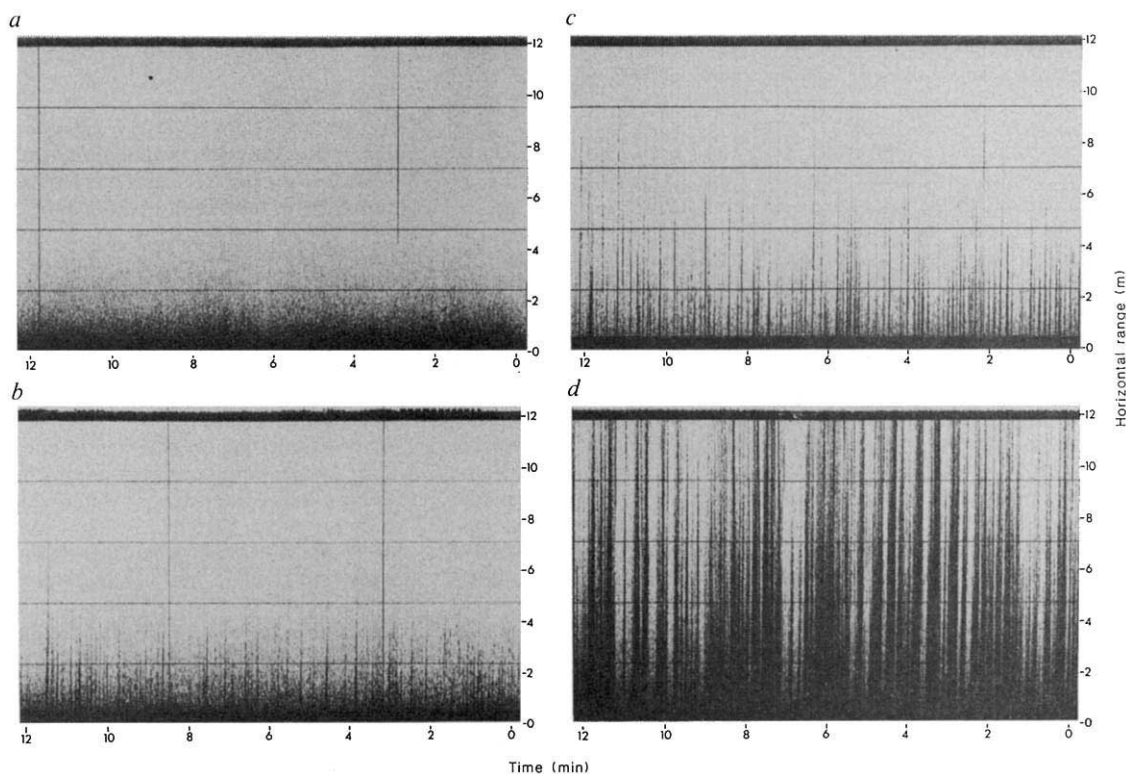


Fig. 1 Sketch of the towed catamaran, spar and sensors.

Figure 2 shows examples of sonograph records made in up-wind tows at different wind speeds. There is little or no scattered sound received in calm conditions (that which does exist may come from biological targets), but in winds of 3 m s^{-1} or more, when waves are breaking, sonar targets are seen which appear to approach the sonar at a speed roughly corresponding to the towing speed. (A tilt is apparent if Fig. 2 is viewed from the bottom.) These returns are from bubble clouds formed by breaking waves²⁻⁵. The sudden onset of a strongly reflecting bubble cloud within the recorded sonar range, 12 m, is rare; most of the clouds persist for tens of seconds⁴⁻⁶. The clouds are composed mainly of very small bubbles^{7,8} which rise very slowly, being almost passive in the wave- and turbulence-dominated near-surface zone³.

Data were selected from tows made within 30° of the wind direction using the sonar returns from 3.38 m ahead of the spar. A zero level and a noise level for the sonar were established by plotting the mean recorded intensity data values and their standard deviations against wind speed to provide estimates at zero wind speed. The zero wind mean plus n standard deviations (the 'noise threshold') have been subtracted from other records to establish 'signals', where $n=2$ or 6. Comparison with earlier work⁶ shows that the noise threshold $n=2$ corresponds to a scattering cross-section for bubble clouds, M_v , of $(1 \pm 0.6) \times 10^{-3} \text{ m}^{-1}$ at 248 kHz whilst that at $n=6$ corresponds to about $M_v = 3 \times 10^{-1} \text{ m}^{-1}$ at 248 kHz. The 'signal' is recognized only when it exceeds these noise thresholds. Scattering becomes more frequent and intense as the wind increases. The number of bubble clouds per wave increases with wind speed from values of about 0.1 (0.01) at 4 m s^{-1} to 0.65 (0.55) at 12 m s^{-1} for $n=2$ (or 6) respectively. The ratio of the length of bubble clouds to the wavelength also increases from about 0.035 (0.015) at 4 m s^{-1} to 0.21 (0.17) at 14 m s^{-1} . In consequence the fraction of record in which bubble clouds are detected by the sonar increases rapidly with wind speed (as can be seen in Fig. 2) approaching a total 'stratus' of bubbles at the depth of the sonar target scattering volume, approximately 3.8 m, at wind speeds of about 15 m s^{-1} . (This stratus scatters sound and, in particular, reduces

Fig. 2 The sonograph records of range ahead of the spar plotted against time (increasing to the left) at wind speeds of *a*, 0; *b*, 5; *c*, 11; and *d*, 14 m s^{-1} . At low wind speeds the slightly tilted bands of reflections from bubble clouds being approached by the spar are absent. The vertical lines in *a* and *b* are time marks. The horizontal lines and the band at 12-m range are range marks.



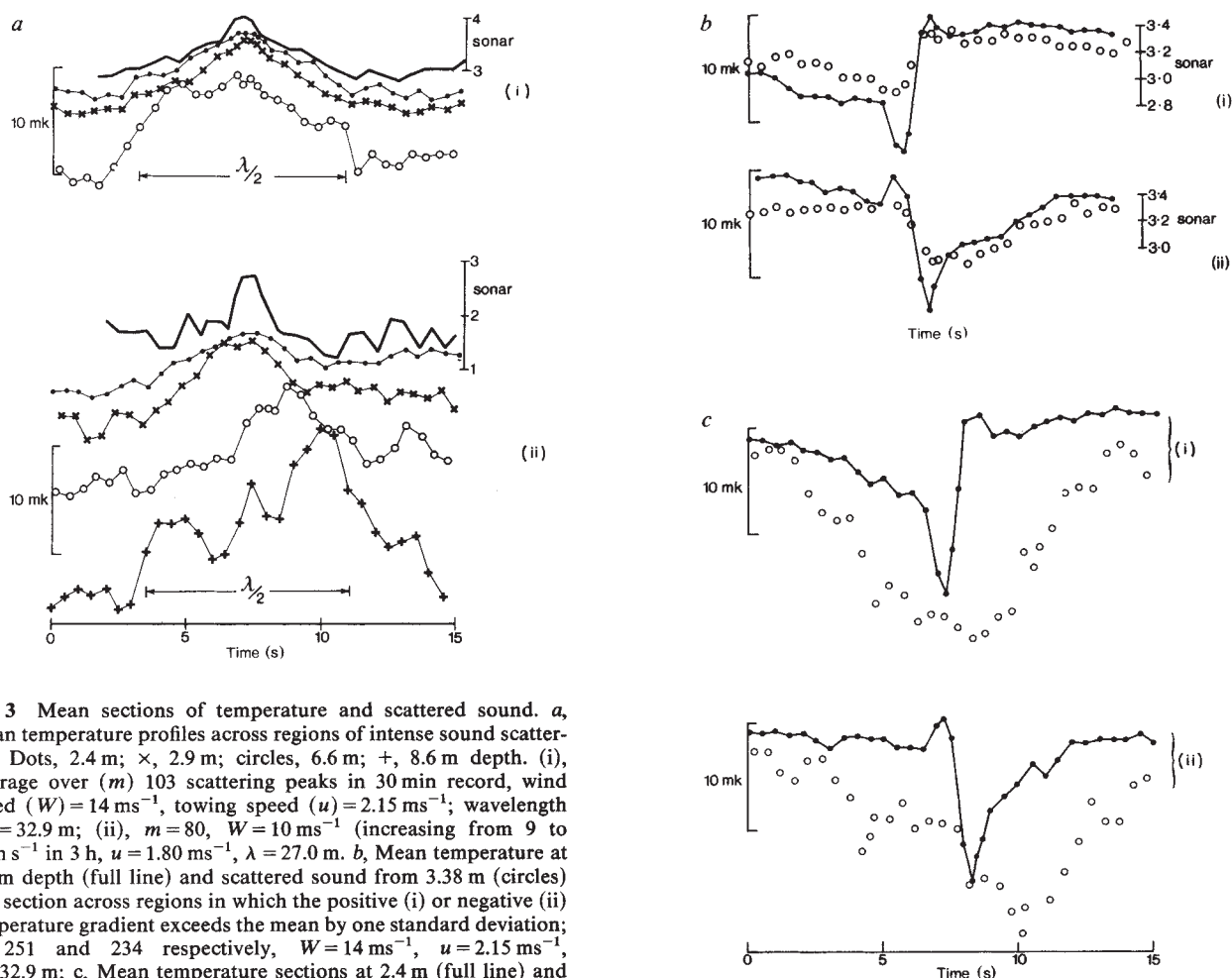


Fig. 3 Mean sections of temperature and scattered sound. *a*, Mean temperature profiles across regions of intense sound scattering. Dots, 2.4 m; $\times$, 2.9 m; circles, 6.6 m; +, 8.6 m depth. (i), Average over (m) 103 scattering peaks in 30 min record, wind speed (W) = 14 ms^{-1} , towing speed (u) = 2.15 ms^{-1} ; wavelength (λ) = 32.9 m; (ii), $m = 80$, $W = 10 \text{ ms}^{-1}$ (increasing from 9 to 15 ms^{-1} in 3 h, $u = 1.80 \text{ ms}^{-1}$, $\lambda = 27.0 \text{ m}$). *b*, Mean temperature at 2.9-m depth (full line) and scattered sound from 3.38 m (circles) in a section across regions in which the positive (i) or negative (ii) temperature gradient exceeds the mean by one standard deviation; $m = 251$ and 234 respectively, $W = 14 \text{ ms}^{-1}$, $u = 2.15 \text{ ms}^{-1}$, $\lambda = 32.9 \text{ m}$; *c*, Mean temperature sections at 2.4 m (full line) and 6.6 m (circles) depth across regions in which the positive (i) or negative (ii) temperature gradient at 2.4 m exceeds the mean by more than one standard deviation; $m = 189$ and 194 respectively, $W = 14 \text{ ms}^{-1}$, $u = 2.08 \text{ ms}^{-1}$, $\lambda = 42.3 \text{ m}$. The temperature and sonar records have been offset to allow for the tilt of the towed spar and the time-lag before the spar reaches the acoustic scattering zone. The sonar scale is log (digitized record number - (mean + $2 \times$ standard deviation of noise)). The threshold is at $M_v = 1 \times 10^{-3} \text{ m}^{-1}$. One thousand logger units, a sonar scale of 3, corresponds to $M_v = 6 \times 10^{-3} \text{ m}^{-1}$, approximately. The relative temperatures are not known precisely; the greater separation between values on either side the sonar peak in 3*a* however implies a larger gradient there than in the central bubble zone.

the amount of the high-frequency sound produced by breaking waves which propagates down into the underlying ocean⁹.)

In detail, the temperature and sonar data are poorly correlated, but patterns emerge when the data are selectively processed. Figure 3*a* shows the average temperature distributions across about 80 bubble clouds in which the peak scattering exceeds the average recorded in a half-hour period by more than two standard deviations. The sonar peaks are, on average, associated with a temperature maximum, the mean temperature being 1–10 mK higher than that on either side. In fairly steady wind conditions (Fig. 3*a*(i)) the widths of the anomalous temperature and sonar regions appear to be about equal (within the resolution of the measurements), suggesting that the turbulence diffuses heat and bubbles in a similar way; the effective 'eddy diffusion coefficients' have approximately equal magnitudes. In a developing sea (Fig. 3*a*(ii)) the signal is still apparent, but less clear. The temperature rise at depth then occurs later than that near the surface, indicating a tilted structure. The trailing angle of the spar is reduced by, on average, 2° as the catamaran passes the centre of the intense bubble clouds. Figure 3*b* is typical of the average sonar response through regions of large positive or negative temperature gradient. On average the intense positive gradients separate regions of differing temperature, that following the rise being usually some 2–5 mK

higher than the preceding region. The warmer zone has also on average a greater concentration of scatterers, often 2–3 times more intense than in the colder zone. The negative temperature gradients are accompanied by a reduction in scattering, but the temperature recovers to its original value within about 10–15 s. The positive and negative temperature gradients at shallow depth have corresponding changes at greater depth (Fig. 3*c*). The rise in temperature at depth, however, lags that nearer the surface, whilst a temperature fall occurs almost simultaneously at all depths. No accompanying trends in the tilt angle of the spar were detected.

The pattern which emerges is one of relatively warm near-surface water being transported downwards, in association with bubbles, into colder deeper water which is deficient of bubbles. Three processes sketched in Fig. 4 may contribute to this transport pattern. Breaking surface waves produce bubbles and downward moving jets¹⁰. Langmuir circulation¹¹ is known to induce bands of bubble clouds in strong winds^{4,12}, and local temperature maxima¹³. Either might account for the temperature-bubble maxima (Fig. 3*a*). Both processes are associated with enhanced down-wind motion near the surface^{4,11} which would increase the drag on the catamaran as it is towed upwind, reducing its forward speed and leading to the observed small reduction in the trailing angle of the spar. The increased

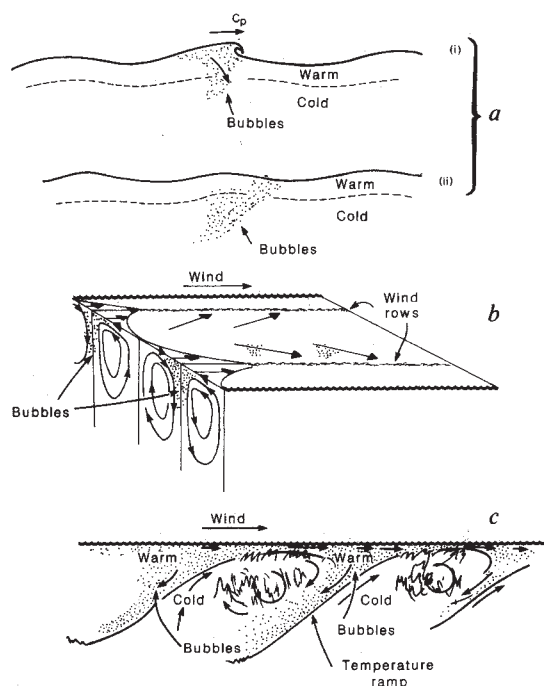


Fig. 4 Sketch showing processes of vertical transport in the upper ocean boundary layer. *a*, Breaking waves: (i), injecting a downward jet of warmer near-surface water and bubbles, (ii), later development. The near-surface water is carried forward by the momentum transferred from the breaking wave. The warm bubble-filled plume is tilted. (See ref. 3, Figs 5c, 6 and 12b for sonographs of tilted bubble clouds.) The lower interface is gravitationally stable; the upper is unstable. *b*, Langmuir circulation. Bubbles formed by waves breaking between the wind rows are carried into the downwelling zones by the circulation. *c*, Eddies and temperature ramps. The instruments were towed upwind, and the ramps are observed as the sensors pass from cold (bubble deficient) water to warm water.

lag of the temperature maxima with depth seen in Fig. 3a(ii) suggests a pronounced presence of shear (see Fig. 4a(ii)) in conditions of increasing wind and more frequent wave breaking. We cannot unambiguously distinguish between these two processes in the present data set, although the frequency of encounter of bands of bubbles formed by Langmuir circulation is expected to be small since the tows were made upwind and so mainly parallel to the wind rows.

The third process is one associated with eddies with axes predominantly aligned cross-wind (in contrast with the downwind alignment of the Langmuir circulations)¹⁴. These lead to temperature ramps^{15,16}, that is, tilted regions of high temperature gradient. The generation mechanism is unclear, possibly an instability of the vertical shear¹⁷. Figure 3b and c are consistent with the pattern sketched in Fig. 4c, the delay in the temperature increase of 2–4 s between 2.4 and 6.6 m corresponding to a ramp slope of 27 to 46° to the horizontal. This is consistent with previous observations¹⁴. Figure 3b and c shows that the temperature ramps are narrow with large changes occurring in 0.5 s (about 1 m horizontally). They are also structured, following and being followed by, thin 'overshoot' regions of lower or higher temperature which often have corresponding lower or higher sonar scattering.

The presence of temperature ramps leads to a skewness, *S*, of the temperature gradient, dependent on the direction of the tow relative to the wind¹. The observed mean value of *S* decreased from 0.55 at 1.4 m depth to 0.38 at 2.7 m, with an average of 0.44, for towing angles θ within 20° of the wind. The average was less, 0.16, for $25^\circ > \theta > 35^\circ$ and also less, but more variable, at $\theta = 0$ below 2.7 m.

The correlation of bubbles and temperature suggests that the turbulence induced by breaking waves may be related to the formation of temperature ramps, the evolution illustrated in Fig. 4a. The upper boundary of the sheared, and tilted, warm bubble cloud is gravitationally unstable through the contributions to density of both heat and bubbles and thus diffuses more rapidly than the lower boundary, a process leading to skewness in the horizontal temperature gradients. If this is so, the symmetrical signals seen in Fig. 3a(i) may be representative of clouds shortly after their generation (for example, at a time very much less than the reciprocal of the vertical shear) whilst the signals associated with temperature ramps, Fig. 3b, are typical of a later stage of evolution. An experiment is presently planned in which these processes can be examined in more detail using a submarine carrying turbulence probes, temperature sensors and upward-looking as well as side-scan sonar to follow the evolution of bubble clouds from their production by breaking waves.

We thank Alan Packwood for his help in developing the towed spar, and Paul Haines for processing the raw data. The data were collected from MV *Sea Searcher* and the work was supported by the Ministry of Defence Procurement Executive under an agreement with the Admiralty Research Establishment, Portland.

Received 4 February; accepted 16 April 1987.

1. Thorpe, S. A. *Nature* **318**, 519–522 (1985).
2. Hall, A. J. & Packwood, A. R. *IOS Internal Rep.* (1987).
3. Thorpe, S. A. *Phil. Trans. R. Soc. A* **304**, 155–210 (1982).
4. Thorpe, S. A. & Hall, A. J. *Cont. Shelf. Res.* **1**, 353–384 (1983).
5. Thorpe, S. A., Hall, A. J., Packwood, A. R. & Stubbs, A. R. *Cont. Shelf. Res.* **4**, 597–607 (1985).
6. Thorpe, S. A. *J. phys. Oceanogr.* **16**, 1462–1478 (1986).
7. Kolovayev, P. A. *Oceanology* **15**, 659–661 (1976).
8. Johnson, B. D. & Cooke, R. C. *J. geophys. Res.* **84**, 3761–3766 (1979).
9. Farmer, D. M. & Lemmon, D. D. *J. phys. Oceanogr.* **14**, 1762–1778 (1984).
10. Su, M., Green, A. W. & Bergin, M. T. in *Gas transfer at Water Surfaces* (eds Brutsaert, W. & Jirka, G. H.) 211–219 (Reidel, Dordrecht, 1984).
11. Leibovich, S. A. *Rev. Fluid Mech.* **15**, 391–427 (1983).
12. Thorpe, S. A. *J. Fluid Mech.* **15**, 391–427 (1984).
13. Thorpe, S. A. & Hall, A. J. *J. Fluid Mech.* **114**, 237–250 (1982).
14. Thorpe, S. A. & Hall, A. J. *J. Fluid Mech.* **101**, 687–703 (1980).
15. Antonia, R. A., Chambers, A. J., Friehe, C. A. & Van Atta, C. W. *J. atmos. Sci.* **36**, 99–108 (1979).
16. Gibson, C. H., Friehe, C. A. & McConnell, S. O. *Phys. Fluids Suppl.* **20**, 156–167 (1977).
17. Thorpe, S. A. *J. geophys. Res.* **83**, 2875–2885 (1978).

The surface windfield over the Antarctic ice sheets

Thomas R. Parish* & David H. Bromwich†

* Department of Atmospheric Science, University of Wyoming, Laramie, Wyoming 82071, USA

† Lared Polar Research Center, The Ohio State University, Columbus, Ohio 43210, USA

The intense radiative cooling of air over the ice slopes of Antarctica generates a surface wind regime that is strongly controlled by topography, and plays a key role in determining the behaviour of the atmosphere and ocean in high southern latitudes^{1–7}. Resultant surface winds are intimately linked to the orientation of the ice terrain (Fig. 1) and display the highest degree of persistence found on Earth. The close coupling between wind and topography allows estimation of the former if the latter is known with some precision. Here we report on time-averaged, near-surface airflow over the Antarctic continent during winter diagnosed from a recent, accurate synthesis of terrain slopes and from estimates of the lower atmospheric temperature structure. The simulated drainage pattern exhibits strong spatial variability with the airflow concentrated into several zones near the coastal margin. These confluence regions are responsible for strong persistent katabatic winds over downstream coastal stretches and are indicative of zones of greatest katabatic potential.

Observational and theoretical studies of gravity-driven slope flows (katabatic winds) are numerous. Ball⁸ points out that the magnitude of terrain-induced pressure gradient is directly proportional to both the steepness of the terrain and the strength of the temperature inversion in the lower atmosphere. Lettau and Schwerdtfeger⁹ refer to this force as the 'sloped-inversion' pressure gradient force to emphasize the interaction between the radiatively forced temperature inversion and the sloping terrain. Other forces which shape the near-surface windfield include the Coriolis force and the frictional force. Large-scale pressure gradients associated with cyclonic storms are probably only of secondary importance because most are situated in the zone of intense thermal contrast between the peripheral ice margins and open ocean and rarely penetrate deep into the Antarctic interior. The strongest winds occur over the steep coastal perimeter where slopes exceed 10^{-2} within the first 150 km inland; they blow almost directly downslope. Winds over the gently sloping continental interior are of more moderate strength and are directed at angles of 40–60° from the fall line. Lettau and Schwerdtfeger⁹ note such interior flows represent a quasi-geostrophic, near-equilibrium wind regime. The term 'inversion' wind has been coined to describe the steady interior surface windfield and to underscore dynamic differences between interior slope flows and the strongly time-dependent coastal 'katabatic' winds.

Although the first-order importance of terrain on the Antarctic surface wind regime has been acknowledged for many years, accurate topographic mapping of the ice contours has only recently been undertaken. Of interest is the series of radio echo soundings carried out by the Scott Polar Research Institute during the past decade. The most refined and up-to-date Antarctic ice topography is presented in Drewry¹⁰, merging the radio echo sounding data with balloon, satellite and oversnow barometric altimetry. The resulting ice contours are thought to be accurate to within approximately ± 30 m, far surpassing the accuracy of previous topographic maps. The importance of refined, precise ice topography maps in understanding the near-surface wind regime over the Antarctic continent cannot be over-emphasized. Parish¹¹ and Parish and Bromwich³ have shown that a realistic diagnosis of the drainage pattern of near-surface winds over the Antarctic continent is possible provided accurate topographic information is available and reasonable estimates of the structure of the temperature inversion are used. The time-averaged streamline maps presented in the above-mentioned works were constructed from the simple steady-state model of Ball⁸ which describes the wind as a balance between the sloped-inversion pressure gradient force, Coriolis force and friction. The steady nature of Antarctic gravity-driven flows over all except the steepest coastal slopes implies a near balance of forces, and the simple Ball model appears to be sufficient for resolving the basic qualitative characteristics of the surface windfield under such conditions. Simulated drainage streamlines over West Antarctica³ and parts of East Antarctica^{11,5} compare favourably with resultant wind directions at interior stations as well as observed sastrugi orientations; sastrugi are streamlined ridges on the snow surface formed by and parallel to the predominant wind.

With the recent completion of a refined topographic map of Antarctica, it is now possible to simulate the time-averaged streamlines of cold air drainage over the entire continent. This is significant because the picture of the surface wind regime will provide *a priori* knowledge of the larger-scale drainage pattern in which a particular site is embedded without the necessity of a costly period of *in situ* measurements. In addition, such a simulation will allow a qualitative estimation of the intensity and persistence of katabatic winds. To diagnose the time-averaged windfield, the entire continent has been digitized to a grid scale of 50 km. This space scale resolves nearly all detail revealed on the Drewry map. Estimates for the temperature inversion over the continent were obtained from the climatologi-

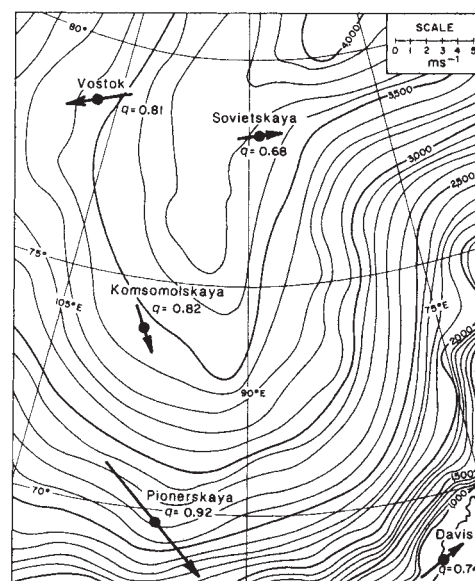


Fig. 1 Annual resultant wind vectors for stations about central ridge in East Antarctica. The symbol q refers to the directional constancy of the wind, the ratio between the vector resultant speed and the mean wind speed. Constancy values near 1 indicate that the airflow is almost exclusively unidirectional.

cal results illustrated in Schwerdtfeger¹². Using the above parameters as the key inputs into the Ball model, equations for the steady motion were solved at each grid point and analyses of the flow directions were then made. To test the sensitivity of the results, a number of runs were carried out, each with a slightly different specification of the inversion strength. The larger-scale drainage features were found to be relatively insensitive to moderate changes in the temperature inversion although the orientation of the drainage patterns do change somewhat in response to the inversion strength differences. This internal consistency was seen in earlier simulations³ and offers testimony to the viability of the simple Ball model in diagnosing qualitative features of Antarctic drainage flows.

Results of the time-averaged near-surface drainage flow pattern from the Ball model are illustrated in Fig. 2. The outstanding feature illustrated in the simulation is the high degree of non-uniformity in the drainage field. The streamlines do not display radially uniform behaviour but rather show a number of clearly defined areas of strong confluence and diffluence. Such patterns have major implications regarding katabatic wind potential and will be discussed in detail later.

Verification of the streamline pattern illustrated in Fig. 2 is understandably difficult owing to the sparse data network over the Antarctic interior. Wind records are available for only about a dozen manned stations in the interior although this data source is currently being supplemented by a number of automatic weather stations (AWS) scattered throughout Antarctica, including several in the continental hinterland. In addition, sastrugi orientation can be used as a proxy indicator of the flow direction. To illustrate the relationship between observed wind direction with the surface streamline simulation in Fig. 2, a composite map has been prepared (Fig. 3) in which resultant wind directions from manned stations and AWS (open arrows) and sastrugi orientations (bold arrows) are plotted relative to a representative number of streamlines from Fig. 2. Previously evaluated sastrugi directions^{3,5,13} have been supplemented by the extensive observations contained within the Japanese Antarctic Research Expedition (JARE) data reports and by data collected during the Australian National Antarctic Research Expeditions (ANARE) traverses. Whenever possible sastrugi directions were

Fig. 2 Time-averaged near-surface wintertime streamlines (heavy lines) of cold air drainage over Antarctica. Thin lines are ice-sheet elevation contours in meters.

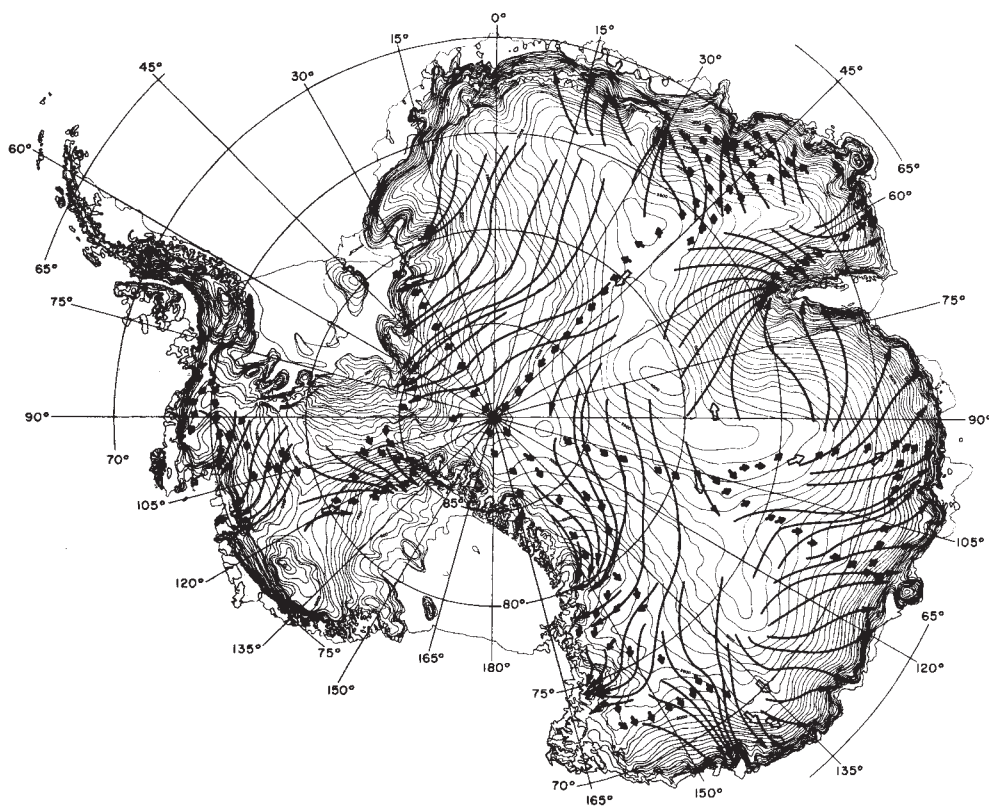
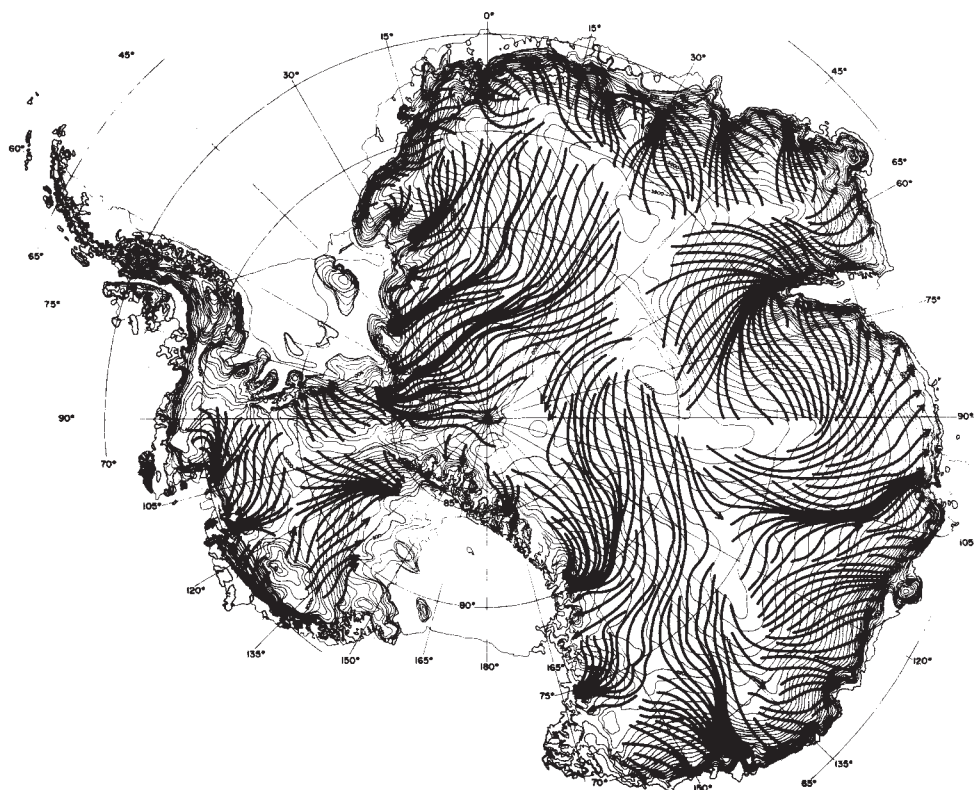


Fig. 3 Composite map of observed annual resultant wind directions (open arrows) and sas-trugi orientations (bold arrows) with streamlines of Fig. 2.



averaged in space and time. Orientations clearly attributable to blizzard-induced snow dunes¹³ have been excluded.

In general, it can be concluded from Fig. 3 that the time-averaged surface streamlines are indeed representative of the large-scale surface windfield. In particular, the wind directions obtained from station data are seen to be highly correlated with the inferred drainage pattern. In certain sections (along the 45° E meridian, for example), sastrugi orientations suggest that the simulated streamlines may be directed too sharply down the fall line. Wind and sastrugi observations taken by E. Mosley-Thompson at 84° S, 43° E during December 1986 indicate that many sastrugi directions over the highest parts of the East Antarctic ice sheet, which were mostly collected during the austral summer, reflect the weaker slope flows of the warm period rather than the stronger, more downslope winter winds which have been modelled. It can be concluded with some confidence that the simulated drainage pattern of Fig. 2 is consistent with observations and therefore is probably a good qualitative estimate of the near-surface time-averaged windfield over the Antarctic ice sheets.

The highly irregular drainage pattern is extremely important in understanding the katabatic regime. A serious limitation to the persistence and intensity of katabatic flow is the supply of negatively-buoyant air upslope in the continental interior. Lettau and Schwerdtfeger⁹ note that even moderate katabatic drainage rapidly exhausts cold air reserves 'as, similarly, the bursting of a dam drains a water reservoir'. The strongly time-dependent behaviour of katabatic flow observed along coastal stretches of Antarctica supports the notion of the importance of such a cold air supply. The episodic high-intensity katabatic surges correspond to the cyclical period of discharging of the cold air reservoir upslope and the subsequent recharging of the katabatic potential. The spatial irregularity of the drainage pattern in the interior of the continent acts to modulate the periodicity of the katabatic regime. Numerical simulations have shown that the available cold air supply reservoir can be significantly enhanced along axes of convergence of air currents¹⁴. In turn, the katabatic regime downslope of this streamline confluence zone becomes intensified and more persistent. Observations have clearly revealed the sensitivity of katabatic flow to interior confluence zones.

The most striking connection is seen in Adelie Land, encompassing the stations of Cape Denison (67.0° S, 142.7° E) and Port Martin (66.8° S, 141.4° E). Cape Denison, the "Home of the Blizzard", was the base camp for the Australasian Antarctic Expedition of 1911-14. Observations of the intense katabatic wind taken during the two-year stay were met with scepticism on the group's return to Australia. Subsequent recalibration of the anemometer as well as the French expedition some 40 years later to nearby Port Martin where similar extreme katabatic wind conditions were encountered have given testimony to the original data set. This coastal section of Adelie Land experiences the strongest and most persistent surface winds on the face of the Earth. Observations at Cape Denison show that the average wind speed is about 19.5 m s⁻¹ (37.9 knots). Mather and Miller¹³ suggest the Adelie Land katabatic winds are about 70% stronger than average. As is clearly revealed in Fig. 2, cold air drainage streamlines upslope of Cape Denison and Port Martin display a strongly confluent pattern. It is believed that this confluence channel with the correspondingly large cold air supply is the fundamental cause of the anomalously intense katabatic regime along the Adelie Land coast.

The association between zones of drainage current convergence in the interior and strong katabatic winds near the coast is further supported by the wind record at Terra Nova Bay (74.9° S, 163.7° E), the region where Scott's Northern Party was forced to winter during 1912. Bromwich and Kurtz¹⁵ reanalysed the party's observations; they conclude that the katabatic winds at Terra Nova Bay are anomalously intense and responsible for the formation of a polynya (open water area) which persists

through each winter. In addition, measurements from an AWS located within the Terra Nova Bay region^{16,17} for the autumn months February, March and April 1984 and 1985 show the average wind speed to be about 17 m s⁻¹ (33 knots), considerably stronger than typical katabatic regimes and comparable to the Adelie Land winds. As is the case for Adelie Land, a zone of drainage streamline confluence can be seen in Fig. 2. Again, the cold air supply reservoir in the interior of the continent is abnormally large, thereby allowing a persistent and intense katabatic airstream.

This work was funded by the NSF. We thank Ian Allison for furnishing both published¹⁸ and unpublished ANARE sastrugi data.

Received 26 January; accepted 16 April 1987.

1. Egger, J. *J. Atmos. Sci.* **42**, 1859-1867 (1985).
2. James, I. N. in *Second Int. Conf. on Southern Hemisphere Meteorology* 117-118 (American Meteorological Society, Boston, 1986).
3. Parish, T. R. & Bromwich, D. H. *Mon. Weath. Rev.* **114**, 849-860 (1986).
4. Polar Research Board, *The Polar Regions and Climatic Change*, Appendix 8-11 (National Academy Press, Washington DC, 1984).
5. Bromwich, D. H. & Kurtz, D. D. *J. Geophys. Res.* **89**, 3561-3572 (1984).
6. Zwally, H. J., Comiso, J. C. & Gordon, A. L. *Antarctic Res. Ser.* **43**, 203-226 (1985).
7. Gill, A. E. *Deep Sea Res.* **20**, 111-140 (1973).
8. Ball, F. K. in *Antarctic Meteorology: Proc. Symp. Melbourne, Australia, 1959*, 9-16 (Pergamon, Oxford, 1960).
9. Lettau, H. H. & Schwerdtfeger, W. *Antarctic J. U.S.* **2**, 155-158 (1967).
10. Drewry, D. J. *Antarctica: Glaciological and Geophysical Folio* (Scott Polar Research Institute, Cambridge, 1983).
11. Parish, T. R. *Mon. Weath. Rev.* **110**, 84-90 (1982).
12. Schwerdtfeger, W. in *World Survey of Climatology* Vol. 14 (ed. Orvig, S.) 253-355 (Elsevier, Amsterdam, 1970).
13. Mather, K. B. & Miller, G. S. *Notes on Topographic Factors Affecting the Surface Wind in Antarctica, with Special Reference to Katabatic Winds: and Bibliography* (Geophysical Institute, University of Alaska, Fairbanks, 1967).
14. Parish, T. R. *Mon. Weath. Rev.* **112**, 545-554 (1984).
15. Bromwich, D. H. & Kurtz, D. D. *Polar Record* **21**, 137-146 (1982).
16. Bromwich, D. H. *Antarctic J. U.S.* **20**, no. 5, 196-198 (1985).
17. Bromwich, D. H. in *Antarctic Climate Res. No. 1* 13-16 (SCAR-ACR, Antarctic Division, Kingston, Tasmania, 1986).
18. Medhurst, T. G. in *ANARE Res. Notes* 28 (ed. Jacka, T. H.) 174-179 (Antarctic Division, Kingston, Tasmania, 1985).

Structure of the glacial thermocline at Little Bahama Bank

Niall C. Slowey* & William B. Curry†

* WHOI/MIT Joint Program in Oceanography, Woods Hole, Massachusetts 02543, USA

† Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

Within the thermocline of the western North Atlantic, bathymetric gradients of hydrographic properties are controlled by the lateral movement of water along isopycnal surfaces¹⁻⁴. This movement is driven by air-sea interaction processes: Ekman pumping and surface heat (buoyancy) flux. By affecting the intensity of these processes, glacial-interglacial climatic change should also affect the bathymetric gradients of properties within the thermocline. Benthic foraminifera live on carbonate bank margins that intersect the thermocline. These foraminifera record the hydrographic gradients in their isotopic shell chemistry, providing the means to reconstruct changes in thermocline structure and circulation. We have measured the $\delta^{18}\text{O}$ compositions of a near-surface planktonic foraminifer, *Globigerinoides ruber*, and a shallow-water benthic foraminifer, *Cibicides floridanus*, in a core recovered from Little Bahama Bank. *Cibicides floridanus* exhibits a glacial-interglacial range in $\delta^{18}\text{O}$ that is 0.5‰ less than *G. ruber*'s, because individuals of this species lived and calcified ~110 m shallower on the thermocline ~18,000 yr ago. After glacio-eustatic changes in thermocline position are accommodated, we estimate that temperature at 540 m palaeodepth cooled by ~2°C. *Globigerinoides ruber* $\delta^{18}\text{O}$ values suggest that surface water temperature at this

location also cooled by $\sim 2^\circ\text{C}$. This uniform cooling of the upper 600 m of the water column is consistent with the observation that surface water where isopycnals outcrop was $\sim 2^\circ\text{C}$ cooler during the glacial maximum⁵⁻⁸.

During January 1982, the RV *Cape Hatteras* recovered piston core CH0182-36 from the southwest margin of Little Bahama Bank, Bahamas. Seismic profiles of the site (26.1797° N, 77.6460° W, 651 m) and the down-core carbonate stratigraphy indicate undisturbed hemi-pelagic deposition in the vicinity of CH0182-36 (refs 9, 10). Sediments are foraminifera-pteropod ooze which appear to be partly lithified at ~ 40 –45 cm depth⁹. Over the depth range of 0–90 cm, four bulk sediment radiocarbon ages^{9,10} (Table 1) and $\delta^{18}\text{O}$ provide chronologic control. We sampled the core at 2-cm intervals for isotopic analyses and present a record with average spacing of 1,000 yr. Oxygen-isotope analyses of *C. floridanus*¹¹ (equivalent to *C. pachyderma* (Rzehak)¹², $>250\ \mu\text{m}$) and *G. ruber* (white variety, 212–250 μm) were made by using standard methods¹³, and our results are presented in δ notation with respect to PDB (Table 1).

The $\delta^{18}\text{O}$ records (Fig. 1) extend from the Late Holocene back through the last glacial period (~ 18 kyr). Ages were assigned to samples by correlating the $\delta^{18}\text{O}$ records of *G. ruber* in CH0182-36 and *Uvigerina* in core V19-30¹⁴ (age model of Shackleton and Pisias¹⁵). Ages obtained by correlation agree reasonably well with the radiocarbon ages (~ 2 kyr mean deviation). Inaccuracies in the V19-30 age model and difficulties in using bulk sediment radiocarbon dates for high resolution stratigraphy¹⁶⁻¹⁸ may cause some differences between the age scales. For this study it is much more important to have closely correlated CH0182-36 with V19-30 than to determine the absolute age of the sediment, because we use V19-30 as both a record of eustatic sea-level and a monitor of the effect of continental ice volume on the $\delta^{18}\text{O}$ of sea water.

There are two notable differences between the planktonic and benthic $\delta^{18}\text{O}$ records. (1) As is typically observed in deep-sea

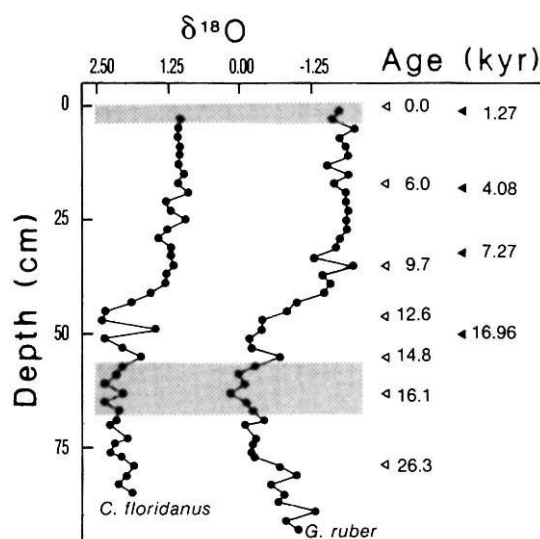


Fig. 1 Oxygen-isotope data for *C. floridanus* and *G. ruber* from core CH0182-36. Correlation points of the $\delta^{18}\text{O}$ records of *G. ruber* in CH0182-36 and *Uvigerina* in V19-30 (refs 14, 15) (open triangles), bulk sediment ^{14}C ages^{9,10} (solid triangles), and both recent Holocene (0–4 cm) and glacial maximum (56–68 cm) $\delta^{18}\text{O}$ data (shaded) are indicated.

sediment cores, $\delta^{18}\text{O}$ values for the benthic *C. floridanus* are always greater than corresponding values for planktonic *G. ruber* because the benthic foraminifer lives in sea water that is cooler (10 – 15°C) than the surface water ($\sim 25^\circ\text{C}$). (2) The magnitude of the glacial-interglacial change of *C. floridanus* (1.1‰) is significantly less than that of *G. ruber* (1.6‰). The differences between the records result from the relative positions of the

Table 1 Visual correlation chronology, ^{14}C ages and oxygen-isotopic data for CH0182-36

Depth (cm)	Age* (kyr)	$\delta^{18}\text{O}_{\text{PDB}}$ (<i>G. ruber</i>)	$\delta^{18}\text{O}_{\text{PDB}}$ (<i>C. floridanus</i>)	Depth (cm)	Age* (kyr)	$\delta^{18}\text{O}_{\text{PDB}}$ (<i>G. ruber</i>)	$\delta^{18}\text{O}_{\text{PDB}}$ (<i>C. floridanus</i>)
0	0.0†			47	12.8	−0.38	2.41
1	0.4	−1.74	—	49	13.3	−0.40	1.47
	1.27 ± 0.06‡			50	16.96 ± 0.29‡		
3	1.1	−1.62	1.04	51	13.8	−0.17	2.37
5	1.8	−2.01	1.07	53	14.3	−0.22	2.06
7	2.5	−1.75	1.08	55	14.8†	−0.74	1.72
9	3.2	−1.86	1.04	57	15.1	−0.27	2.06
11	3.9	−1.90	1.05	59	15.5	0.02	2.15
13	4.6	−1.52	1.06	61	15.8	−0.10	2.37
15	5.3	−1.92	0.98	63	16.1†	0.17	2.01
17	6.0†	−1.64	1.08	65	17.4	−0.12	2.36
18	4.08 ± 0.10‡			67	18.7	−0.23	2.12
19	6.4	−1.86	0.90	69	20.0	−0.44	2.15
21	6.8	−1.85	1.29	70	20.6	−0.09	2.28
23	7.2	−1.90	1.20	73	22.6	−0.29	1.95
25	7.6	−1.87	0.95	74	23.2	−0.24	2.19
27	8.0	−1.87	1.27	76	24.5	−0.21	2.25
29	8.5	−1.75	1.42	77	25.1	−0.27	2.06
31	8.9	−1.69	1.21	78	26.3†		
32	7.27 ± 0.11‡			79	26.4	−0.73	1.84
33	9.3	−1.31	1.21	81	27.7	−1.00	1.97
35	9.7†	−2.00	1.16	83	29.0	−0.55	2.10
37	10.2	−1.44	1.28	85	30.3	−0.79	1.87
39	10.8	−1.58	1.30	87	31.5	−0.67	—
41	11.3	−1.47	1.56	89	32.8	−1.34	—
43	11.8	−1.00	1.89	91	34.1	−0.82	—
45	12.3	−0.82	2.35	93	35.4	−1.03	—
46	12.6†						

* Unless otherwise noted, obtained by linear interpolation between tie points from visual correlation between $\delta^{18}\text{O}$ of *G. ruber* and V19-30 *Uvigerina*^{14,15}.

† Visual correlation tie point.

‡ From bulk sediment ^{14}C dating^{9,10}.

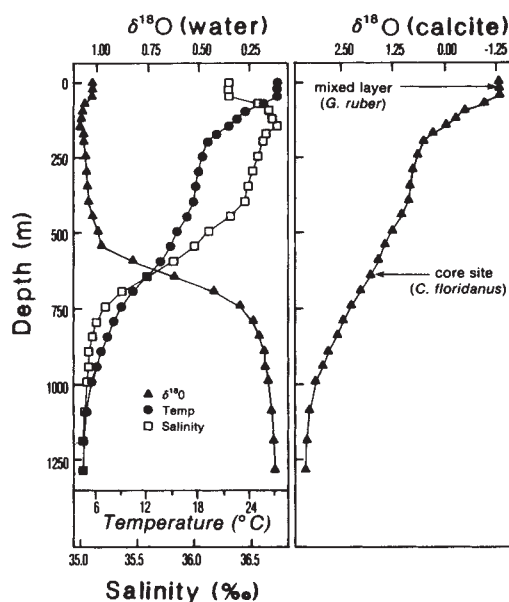


Fig. 2 a, Profiles of temperature, salinity and $\delta^{18}\text{O}$ for the present-day water column near the site of CH0182-36 (25.7833° N, 77.0000° W, RV Researcher cruise ANT80). Values of $\delta^{18}\text{O}$ were calculated from the salinity data by using salinity- $\delta^{18}\text{O}$ relations (R. G. Fairbanks, personal communication) for Sargasso Sea Water and North Atlantic Deep Water. b, Profile of the $\delta^{18}\text{O}$ of equilibrium calcite and the depths at which *G. ruber* and *C. floridanus* live today.

planktonic and benthic foraminifera within the thermocline, and the vertical migration of the thermocline along the bank margin (where *C. floridanus* lives) that accompanies glacio-eustatic sea level changes.

The $\delta^{18}\text{O}$ of calcite deposited in equilibrium with the present sea water at Little Bahama Bank increases with depth because of a large decrease in temperature within the thermocline. Figure 2a shows the measured temperature and salinity (25.7833° N, 77.0000° W, RV Researcher cruise ANT80), and estimated oxygen isotopic composition of sea water ($\delta^{18}\text{O}_w$) as a function of depth as it is today near the site of CH0182-36. Because the principal source of this water is the Gulf Stream recirculation system¹⁹, $\delta^{18}\text{O}_w$ was calculated from $\delta^{18}\text{O}_w$ -salinity relations for Sargasso Sea Water and North Atlantic Deep Water (R. G. Fairbanks personal communication). In the surface waters, salinity decreases from lower to higher latitudes, and higher salinities correlate with enriched $\delta^{18}\text{O}_w$ values because evaporation increases salinity while fractionating H_2^{16}O from H_2^{18}O ^{20,21}. Both salinity and $\delta^{18}\text{O}_w$ values decrease with depth at this location because deeper thermocline waters are supplied along isopycnals from the ocean surface at higher latitudes¹⁻⁴. Despite this decrease in $\delta^{18}\text{O}_w$, the $\delta^{18}\text{O}$ of calcite precipitated in equilibrium with the sea water shows a consistent increase with depth in the water column (Fig. 2b) because of the overwhelming effect of the large temperature decrease^{22,23}.

The $\delta^{18}\text{O}$ difference between *G. ruber* and *C. floridanus* reflects this bathymetric gradient. *Globigerinoides ruber* live and deposit their tests within the mixed layer²⁴⁻³⁰, so their equilibrium $\delta^{18}\text{O}$ composition should equal -1.3‰. The mean value observed for the interval 0.4-1.1 kyr is -1.7‰. The observed difference between our recent Holocene average and the calculated equilibrium $\delta^{18}\text{O}$ value is -0.4‰, which agrees with previously reported disequilibria of -0.25-0.5‰²⁸⁻³⁰. Much of this difference has been attributed to the effects of symbiotic algae²⁹. Sediment-trap studies in the Sargasso Sea indicate that the flux of *G. ruber* to the sea floor is relatively constant throughout the

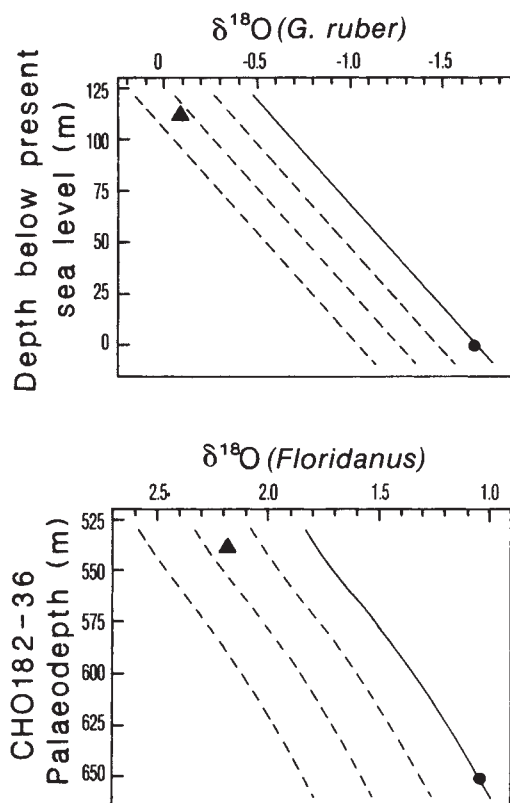


Fig. 3 The average measured $\delta^{18}\text{O}$ of recent Holocene foraminifera are represented by circles. If water temperatures near the surface and within the thermocline have remained the same from the last glacial maximum through today, then the solid lines trace the $\delta^{18}\text{O}$ of foraminifera that would be expected as ice volume and sea level changed; dashed lines trace the $\delta^{18}\text{O}$ of foraminifera if water temperatures were decreased by 1°C increments. For a given y-axis value, heavier $\delta^{18}\text{O}$ values indicate cooler water temperatures and the dashed lines provide a scale to calibrate the temperature of water in which foraminifera lived. The averaged measured $\delta^{18}\text{O}$ of glacial foraminifera (triangles) are heavier than expected, indicating that glacial water temperatures were cooler than today at (a) the surface and (b) 540 m palaeodepth.

year^{31,32}. We assume that this same seasonal pattern of *G. ruber* production existed during the glacial maximum, so that the $\delta^{18}\text{O}$ of *G. ruber* reflects the average annual surface water palaeotemperature. For individual *Cibicidoides* species, the offset from equilibrium has been reported to range between -0.64 and -1.0‰³³⁻³⁷. Our most recent Holocene value is 1.0‰ and expected equilibrium at 651-m water depth is 1.8‰, indicating that vital effects of *C. floridanus* fall within the range of previously investigated *Cibicidoides* species. We assume that the vital effects remain constant through time so that foraminiferal and equilibrium $\delta^{18}\text{O}$ values have the same spatial and temporal gradients.

Figure 3 shows the average measured interglacial and glacial maximum $\delta^{18}\text{O}$ values of *G. ruber* and *C. floridanus* along with the glacial foraminifera $\delta^{18}\text{O}$ values that would be expected if the thermocline maintained its present depth structure while the sea level changed. Also shown are the $\delta^{18}\text{O}$ values that would exist if sea-water temperatures at this location decreased by increments of 1°C. Palaeodepths and past sea-level are based on the history of eustatic sea-level inferred from the $\delta^{18}\text{O}$ record of benthic foraminifera in V19-30^{14,15} (corrected for effects of deep-water temperature change³⁸). Eustatic sea-level changes were calculated by assuming that a 0.1‰ increase in $\delta^{18}\text{O}$ corresponds to a 10 m fall in sea-level^{38,39}. The expected $\delta^{18}\text{O}$ values were generated by the palaeotemperature equation^{22,23} and the

data in Fig. 2 (after accounting for ice volume effects as indicated by the V19-30 record).

For the surface-dwelling planktonic foraminifer, *G. ruber*, the expected $\delta^{18}\text{O}$ composition follows a simple response to sea-level change because its position within the water column does not change with respect to sea-level. The benthic foraminifer, *C. floridanus*, has a more complicated response to sea-level change because its $\delta^{18}\text{O}$ composition is affected not only by enrichment and depletion that correspond to ice volume growth and decay, but also by changes in its position within the thermocline as sea level changes. Thus whereas the isotopic composition of *G. ruber* is affected mostly by changes in ice volume and sea surface temperature, *C. floridanus* is affected by ice volume, migration of the thermocline and changes in the temperature structure within the thermocline.

The average glacial $\delta^{18}\text{O}$ compositions for both *G. ruber* and *C. floridanus* show that near-surface and mid-thermocline water temperatures were colder by about 2.3 and 1.7 °C, respectively, during the last glacial maximum than today. If only differences in ice volume and eustatic sea level change affected the $\delta^{18}\text{O}$ of *C. floridanus*, then its glacial $\delta^{18}\text{O}$ value should be 1.8‰ during the lowest sea level (sea-level change -110 m, CH0182-36 palaeodepth ~540 m) (Fig. 3b). The predicted net difference from today of 0.7‰ reflects the combined effects of a 1.1‰ change due to greater ice volume and a -0.4‰ change due to calcification shallower in the thermocline at warmer temperatures (see Fig. 2b). This predicted glacial $\delta^{18}\text{O}$ value is 0.4‰ less than the measured glacial value, indicating that the thermocline at that paleodepth (540 m) was about 1.7 °C cooler than at 540 m depth in the modern ocean. Similarly, if glacial near-surface waters had the same temperature as today, then the predicted $\delta^{18}\text{O}$ value of *G. ruber* would be -0.6‰ during the glacial maximum, 1.1‰ greater than today due to greater ice volume (Fig. 3a). This value is 0.5‰ less than our measured glacial value for *G. ruber*, indicating that surface water temperature cooled by 2.3 °C. Precisely how variations in evaporation and precipitation affected the $\delta^{18}\text{O}$ of the water during glacial times is not known. Our glacial-interglacial surface temperature decrease agrees closely with faunal assemblage estimates (about -2 °C) for the same region^{5-8,40,41}, implying that any such effects on near-surface water are small.

We have assumed that eustatic and local sea levels are the same. Because the site of CH0182-36 is more than 200 km from the continental margin of eastern Florida, we suspect that there is little difference between eustatic and local sea levels. In any case, using a simple Airy model of isostatic compensation, we estimate that the difference between changes of eustatic and local sea level was at most 34 m. Consequently, our estimate of the glacial-interglacial change in mid-thermocline temperatures may be in error by up to 0.4 °C because of isostatic effects. The effects of even this maximum possible difference in sea level change would hardly be detected because the mid-thermocline temperature estimate from the $\delta^{18}\text{O}$ of *C. floridanus* is not more precise than about ± 0.3 °C.

Our temperature estimates require that the effect of glacial ice volume on $\delta^{18}\text{O}_w$ has been correctly determined by Chappel and Shackleton³⁸. However, the $\delta^{18}\text{O}$ difference between *G. ruber* and *C. floridanus* can be used to reconstruct the past vertical structure (i.e. temperature gradient) of the water column at Little Bahama Bank without this stipulation. As illustrated in Fig. 3, if the same thermocline structure existed during the glacial climate, then the glacial-interglacial range for *C. floridanus* should be 0.4‰ less than for *G. ruber*. Our measurements show that the $\Delta\delta^{18}\text{O}$ value for *C. floridanus* is 0.5‰ less than that of *G. ruber*. Thus no changes in thermocline structure are necessary to accommodate the differences between the surface water and mid-thermocline isotopic records. If our palaeotemperature estimates are correct, glacial cooling of the surface ocean affected at least the upper 600 m of the water column.

These results are consistent with inferences drawn from the simple ventilated thermocline model of Luyten *et al.*^{3,42}. In this model, the structure of the thermocline reflects the latitudes at which isopycnal surfaces outcrop at the ocean surface and the strength of wind-driven Ekman pumping. During the last glacial maximum, stronger trade winds^{5,8,43-49} and westerlies^{8,50} increased Ekman pumping, and isopycnal surfaces outcropped at lower latitudes because of decreased surface temperature⁵⁻⁸ and southward migration of the oceanic polar front^{5,6,51}. Consistent with the model of Luyten *et al.*^{3,41}, our data indicate that the thermocline structure at the latitude, longitude and palaeodepth (~540 m) of CH0182-36 was relatively insensitive to the glacial-interglacial changes in either Ekman pumping or isopycnal outcrop location. The ~2 °C glacial cooling of the upper water column indicated by our data is consistent with the observation that much of the ocean surface where the isopycnal surfaces outcropped was ~2 °C cooler⁵⁻⁸.

We thank D. Graham, L. Keigwin and G. P. Lohmann for critical reviews; S. Burns and A. C. Neumann for permission to sample CH0182-36; D. Olson for hydrographic data; W. A. Berggren for providing the correct identification of the shallow-water benthic *C. floridanus*; and, especially, D. Ostermann and J. Cole for help with foraminifera identification and oxygen-isotope analysis. Supported by a graduate fellowship from the WHOI/MIT Joint Program in Oceanography to N.C.S. and NSF grants to W.B.C. Collection and storage of CH0182-36 was supported by an NSF grant to A. C. Neumann and M. R. Boardman.

Received 22 September 1986; accepted 6 April 1987.

- Islén, C. O'D. *Trans. Am. geophys. Un.* **20**, 414-417 (1939).
- Jenkins, W. J. *J. mar. Res.* **38**, 533-569 (1980).
- Luyten, J., Pedlosky, J. & Stommel, H. *J. Phys. Oceanogr.* **13**, 292-309 (1983).
- Luyten, J. & Stommel, H. *J. J. phys. Oceanogr.* **16**, 1551-1560 (1986).
- CLIMAP Project Members *Science* **191**, 1131-1137 (1976).
- McIntyre, A. *et al. Geol. Soc. Am. Mem.* **145**, 43-76 (1976).
- CLIMAP Project Members *Geol. Soc. Am. Map Chart Ser.* **36** (1981).
- Manabe, S. & Broccoli, A. J. *J. geophys. Res.* **90**, 2167-2190 (1985).
- Burns, S. J. M. S. thesis Univ. North Carolina (1983).
- Burns, S. J. & Neumann, A. C. *Marine Geology* (Submitted).
- Poag, C. W. *Ecologic Atlas of Benthic Foraminifera of the Gulf of Mexico* (Hutchison Ross, Stroudsburg, 1981).
- van Morkhoven, F. P. C. M., Berggren, W. A. & Edwards, A. S. *Cenozoic Cosmopolitan Deep-water Benthic Foraminifera* (Elf Aquitaine, Pau, 1986).
- Curry, W. B. & Lohmann, G. P. *Quat. Res.* **18**, 218-235 (1982).
- Shackleton, N. J., Imbrie, J. & Hall, M. A. *Earth planet. Sci. Lett.* **65**, 233-244 (1983).
- Shackleton, N. J. & Pisias, N. G. *Geophys. Monog.* **32**, 303-317 (1985).
- Keigwin, L. D., Corliss, B. H., Druffel, E. R. M. & Laine, E. P. *Quat. Res.* **22**, 383-386 (1984).
- Berger, W. H., Killingley, J. S. & Vincent, E. *Nature* **314**, 156-158 (1985).
- Duplessy, J. C., Arnold, M., Maurice, P., Bard, E., Duprat, J. & Moyes, J. *Nature* **320**, 350-352 (1986).
- Olson, D. B., Schott, F. A., Zantopp, R. J. & Leaman, K. D. *J. phys. Oceanogr.* **14**, 1470-1487 (1984).
- Epstein, S. & Mayeda, T. *Geochim. cosmochim. Acta* **4**, 213-224 (1953).
- Broecker, W. S. *Quat. Res.* **26**, 121-134 (1986).
- Epstein, S., Buchsbaum, R., Lowenstam, H. & Urey, H. C. *Bull. geol. Soc. Am.* **64**, 1315-1326 (1953).
- Craig, H. in *Proc. Spoleto Conf. on Stable Isotopes in Oceanographic Studies and Paleotemperatures* Vol. 3 (ed. Tongiorgi, E.) (1965).
- Emiliani, C. *Am. J. Sci.* **252**, 149-158 (1954).
- Jones, J. I. *Micropaleont.* **13**, 489-501 (1967).
- Lidz, B., Kehm, A. & Miller, H. *Nature* **217**, 245-247 (1968).
- Berger, W. H. *Deep Sea Res.* **16**, 1-24 (1969).
- Shackleton, N. J. & Vincent, E. *Mar. Micropaleont.* **3**, 1-13 (1978).
- Fairbanks, R. G., Wiebe, P. H. & Be, A. W. H. *Science* **207**, 61-63 (1980).
- Curry, W. B. & Matthews, R. K. *Mar. Micropaleont.* **6**, 327-337 (1981).
- Deuser, W. G. *Deep Sea Res.* (in the press).
- Deuser, W. G. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **33**, 103-127 (1981).
- Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* **3**, 39-55 (1973).
- Shackleton, N. J. in *Fate of Fossil Fuel CO₂* (eds Andersen, N. R. & Malahoff, A.) 401-427 (Plenum, New York, 1977).
- Woodruff, F., Savin, S. M. & Douglas, R. G. *Mar. Micropaleont.* **5**, 3-11 (1980).
- Belanger, P. E., Curry, W. B. & Matthews, R. K. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **33**, 205-220 (1981).
- Graham, D. W., Corliss, B. H., Bender, M. L. & Keigwin, L. D. *Mar. Micropaleont.* **6**, 483-497 (1981).
- Chappell, J. & Shackleton, N. J. *Nature* **324**, 137-140 (1986).
- Fairbanks, R. G. & Matthews, R. K. *Quat. Res.* **10**, 181-196 (1978).
- Hunter, G. E. M. S. thesis Miami Univ. (1984).
- Boardman, M. R., Neumann, A. C., Baker, P. A., Dulin, L. A., Kenter, R. J., Hunter, G. E. & Kiefer, K. B. *Geology* **14**, 28-31 (1986).
- Luyten, J., Pedlosky, J. & Stommel, H. *Clim. Change* **5**, 183-191 (1983).
- Hays, J. D. & Perruzza, A. *Quat. Res.* **2**, 355-362 (1972).

44. Parkin, D. W. & Shackleton, N. J. *Nature* **245**, 455-457 (1973).
 45. Parkin, D. W. *Proc. R. Soc. Lond. A* **337**, 73-100 (1974).
 46. Parkin, D. W. & Padgham, R. C. *Proc. R. Soc. Lond. A* **346**, 245-260 (1975).
 47. Gates, W. L. *Science* **191**, 1138-1144 (1976).
 48. Gardner, J. V. & Hays, J. D. *Geol. Soc. Am. Mem.* **145**, 221-246 (1976).
 49. Kolla, V., Biscaye, P. E. & Hanley, A. *Quat. Res.* **11**, 261-277 (1979).
 50. Kutzbach, J. E. & Guetter, P. J. *J. Atmos. Sci.* **43**, 1726-1759 (1986).
 51. Ruddiman, W. F. & McIntyre, A. *Quat. Res.* **3**, 117-130 (1973).

Tree cellulose $^{13}\text{C}/^{12}\text{C}$ isotope ratios and climatic change

Minze Stuiver & Thomas F. Braziunas

Department of Geological Sciences and Quaternary Research Center, University of Washington, Seattle, Washington 98195, USA

Because tree cellulose is formed from carbon acquired through photosynthesis, its $^{13}\text{C}/^{12}\text{C}$ isotope ratio reflects that of atmospheric CO_2 . Fractionation changes related to photosynthetic reactions also influence $^{13}\text{C}/^{12}\text{C}$ values, and an offset between atmospheric and tree cellulose isotope ratios results. Here, we measured the pre-1850 cellulose $^{13}\text{C}/^{12}\text{C}$ isotope ratios of 19 North American coniferous trees and found a strong dependency of the ratios on latitude. Relative humidity and perhaps temperature appear to be important for this relationship. Assuming climate change to be the dominant factor influencing pre-1850 tree cellulose isotope ratios, we compare a 2,000-yr-long record of $^{13}\text{C}/^{12}\text{C}$ isotope change with existing records of ice acidity and temperature¹. When accepting a lag in $^{13}\text{C}/^{12}\text{C}$ response to climatic change of 70 to 90 yr, correlation coefficients of 0.8 are found for the 1100-1850 interval.

The magnitude of the $^{13}\text{C}/^{12}\text{C}$ offset between tree cellulose and atmospheric CO_2 depends on many environmental variables (precipitation, temperature, physiology, light intensity, nutrients, and relative humidity)² and as a result contradictory evidence for a tree $\delta^{13}\text{C}$ -temperature ($\delta^{13}\text{C} = [(^{13}\text{C}/^{12}\text{C}_{\text{sample}} - ^{13}\text{C}/^{12}\text{C}_{\text{PDB}}) / (^{13}\text{C}/^{12}\text{C}_{\text{PDB}})] \times 1,000$ per mil) link exists³⁻⁵. Some apparent relationships, however, are those given by Pearman *et al.*⁶ for *Arthrotaxis selaginoides* ($\alpha = 0.48 \pm 0.12$ and 0.24 ± 0.11 per $^{\circ}\text{C}$), by Wilson and Grinstead⁷ for *Pinus Radiata* ($\alpha = 0.2$ per $^{\circ}\text{C}$), by Freyer and Belacy⁸ for *Pinus Silvestris* ($\alpha = 0.18$ per $^{\circ}\text{C}$) and by Tans⁹ for *Quercus* ($\alpha = 0.27$ to 0.39 per $^{\circ}\text{C}$). A negative coefficient of -0.27 per $^{\circ}\text{C}$ December temperature has been reported for *Juniperus*¹⁰. *Juniper* leaf cellulose, however, has a positive temperature coefficient of 0.13 per $^{\circ}\text{C}$ for average spring temperatures between 6° and 15°C ¹¹.

Relative humidity influences plant $\delta^{13}\text{C}$ also. A lowering of relative humidity by about 48% increased $\delta^{13}\text{C}$ values of two C-3 grasses by 0.7 to 2.7‰ (equivalent to -0.015 to -0.048 ‰/per cent relative humidity). Similar results were obtained for chickpea (*Cicer arietinum* L.)^{12,13}. Lower relative humidity results in an increased ratio of CO_2 assimilation to transpiration.

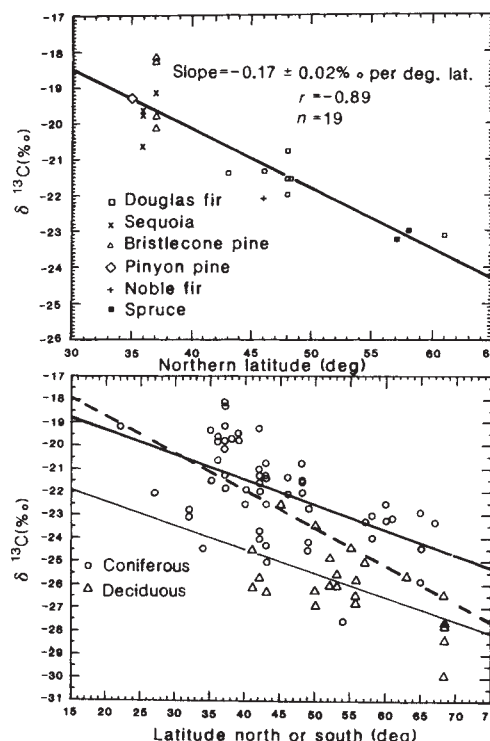


Fig. 1 a, Upper section: latitude dependency of pre-1850 average decadal $\delta^{13}\text{C}$ of western North American conifers. b, Lower section: latitude dependency of $\delta^{13}\text{C}$ of coniferous (heavy solid line) and deciduous (light solid line) trees, indicating a 3.0 per mil off-set between groupings. Dashed line is the best fit to all data points.

If indeed the positive relationship between tree cellulose $\delta^{13}\text{C}(\delta_i)$ and temperature found above for a limited number of trees plays a role, the δ_i values of tree cellulose should be higher at lower latitudes. Such a relationship with latitude is observed for our pre-AD 1850 average cellulose δ_i values of 19 North American coniferous trees (Fig. 1a). The correlation coefficient $r = 0.89$ ($n = 19$) indicates a significance level p of better than 0.1% (the probability that the variables are not correlated is less than 0.1%). In Fig. 1b we have added δ_i values of coniferous (squares) and deciduous (triangles) trees of several other studies^{6,8,14-26}. The relationship still holds, although δ_i values scatter more. The latter is not surprising because for Fig. 1a all materials are cellulose, whereas part of the Fig. 1b additions are whole wood or 'de Vries' treated wood²⁷. Also, some of the additional data are for fairly recent (twentieth century) samples. These samples could have been subjected to δ_i lowering due to fossil fuel combustion. Juvenile effects² may influence some of the Fig. 1b materials, whereas the increased number of species also plays a role. Yet, despite these additional sources of δ_i ,

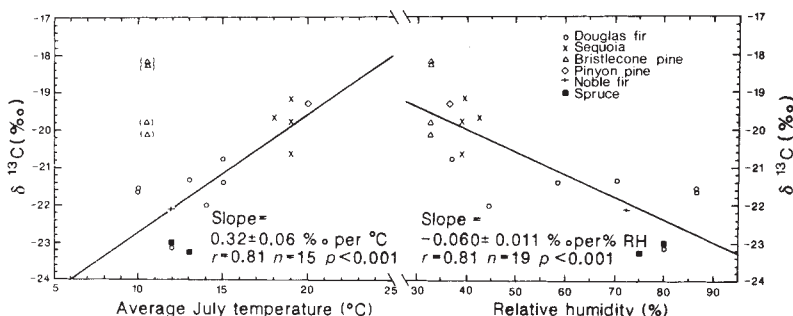
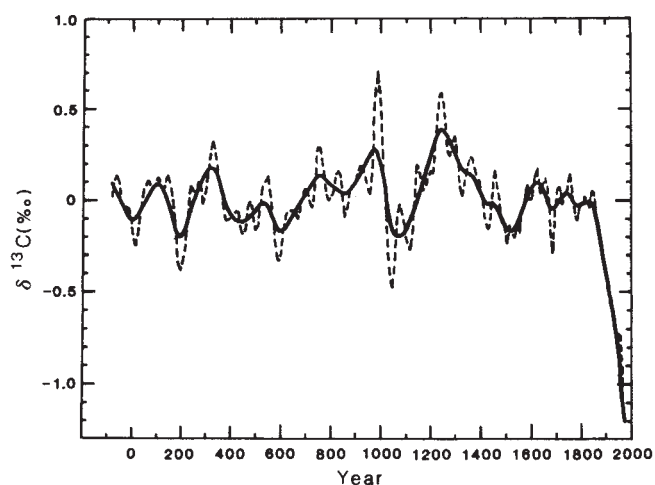


Fig. 2 Average summer temperature- $\delta^{13}\text{C}$ (left) and relative humidity- $\delta^{13}\text{C}$ relationship of 19 trees from North America. The temperature- $\delta^{13}\text{C}$ regression line does not take into account the Bristlecone Pine values (see text).

Fig. 3 Residual (pre-1850 averages deducted) decadal $\delta^{13}\text{C}$ of conifers of the west coast of North America. A cubic spline (dashed curve) takes into account the standard deviations in the residual $\delta^{13}\text{C}$ values. Doubling the standard deviations results in the smoothed spline (heavy solid curve).



variance, the δ_i -latitude significance level is better than 0.3% for the coniferous (slope = $-0.11 \pm 0.02\%$ per degree latitude) and deciduous (slope = $-0.10 \pm 0.03\%$ per degree latitude) subsets. The deciduous and coniferous δ_i values differ, on average, by 3.0%. Combining all tree δ_i data, and neglecting the differences between deciduous and coniferous species, leads to the dashed line in Fig. 1b ($r = 0.67$, $n = 78$, $p < 0.001$, slope = $-0.16 \pm 0.02\%$ per degree latitude).

Whereas a temperature influence of δ_i is an obvious candidate for a latitude- δ_i relationship, such a relationship could also result from relative humidity changes^{12,13,28}. Our mid-latitude North American conifers were collected for their long records, and these trees (Bristlecone Pine, Sequoia) grew at high elevations where low relative humidities are prevalent. If humidity is the dominant factor, one expects higher δ_i at intermediate to low latitudes. As lower δ_i values would be expected for high humidity (sub)tropical forests, a δ_i maximum would result. Such a maximum agrees with the relatively low δ_i value of -26% measured by us for an unidentified tree of Costa Rica, and with the δ_i of -28% found for Amazonian wood (P. D. Quay, personal communication). It should be noted, however, that tropical δ_i values of modern wood also may be lowered by extensive recycling of biospheric CO_2 , or anthropogenic post-1850 change. Similarly, it can be argued that a temperature-related δ_i maximum at intermediate latitudes is also a possibility because a temperature optimum for photosynthesis exists which in warmer climates could result in a reverse temperature effect for certain species²⁹.

We do not find significant correlations between δ_i and average summer temperatures for either our complete set of North American trees, nor all our measured conifers (19 from North America, 2 from South America and 4 from New Zealand). But without the highly stressed Bristlecone Pines (growing at $\sim 3,200\text{-m}$ elevation) we do find a δ_i -temperature correlation (Fig. 2, $\alpha = 0.32 \pm 0.06\%$ per $^\circ\text{C}$, $r = 0.81$, $n = 15$, $p < 0.001$) for our remaining North American trees, as well as our combined conifer data set ($\alpha = 0.21 \pm 0.08\%$ per $^\circ\text{C}$, $r = 0.51$, $n = 21$, $p = 0.02$). For these calculations the average summer temperatures (which usually parallel growing season temperatures) were estimated from weather station records (using the nearest stations with altitude close to that of tree growth and applying a lapse rate of $-6\text{ }^\circ\text{C km}^{-1}$ elevation change). Assuming relative humidity to be the controlling factor we find a highly significant correlation between δ_i and relative humidity (Fig. 2, $\alpha = -0.06 \pm 0.01\%$ per cent relative humidity, $r = 0.81$, $n = 19$, $p < 0.001$). Average July daytime relative humidity was derived from records of weather stations in the proximity of the tree growth sites. Corrections for elevation differences were made by applying standard pressure-temperature diagrams³⁰.

We were not able to locate proper temperature, humidity and elevation data for many of the samples not measured in Seattle. But for those samples for which we have approximate elevations we find that the sampled low-latitude conifers grew at higher elevations. This results in a significant correlation between δ_i and elevation ($r = 0.60$, $n = 44$, $p < 0.001$). For deciduous trees a relationship between δ_i and elevation ($n = 20$) is absent. And when the subset of conifers growing below 700-m elevation is selected, the δ_i -latitude relationship is still preserved ($r = 0.55$, $n = 13$, $p = 0.005$). Evidently elevation is not a primary causative factor for δ_i change. Relative humidity and temperature appear to be the controlling factors, with the former the leading candidate.

The empirically derived latitude relationship of tree cellulose (Fig. 1) and the cited instances of δ_i change related to relative humidity (and temperature) in specific plants show that long-term records of δ_i contain palaeoclimatic information. Figure 3 gives such a long-term record. It was derived from the residual δ_i values of 18 North American coniferous trees (the record of the nineteenth tree was too short) by deducting the average pre-1850 δ_i values of the individual trees. Not all records are two thousand years long. Three to five trees cover the first millenium, three to nine the first half of the current millenium, and ten to seventeen the post-1500 interval. The dashed curve was obtained by fitting a cubic spline $\delta_i = g(x)$ to the $n + 1$ data points such that

$$\sum_{i=0}^{i=n} \left(\frac{g(x_i) - \delta_{ti}}{\sigma_{ti}} \right)^2 = n + 1$$

where x_i is the decade, σ_{ti} the standard deviation of the residual δ_i values, and δ_{ti} the corresponding δ_i value for the decade. More relevant for our discussion, however, is the smoothed heavy solid spline which was obtained by doubling σ_{ti} . The pronounced post-1850 δ_i decline, which is caused by fossil fuel and deforestation CO_2 emissions to the atmosphere²⁶, will be discussed elsewhere. Our focus here is on the pre-1850 δ_i variability.

The Fig. 4 temperature index cubic spline is representative of some of the general temperature trends in the Northern Hemisphere¹. The temperature index is based on a combination of Central England temperatures, California (White Mountains) tree ring width indices, and Greenland $^{18}\text{O}/^{16}\text{O}$ isotope ratios. The Northern Hemisphere (NH) index contains the 'Little Ice Age' interval (AD 1350–1900) and the 'Medieval Warm' interval (AD 1000 to 1300). There is an increasing uncertainty in the NH temperature index backward in time (for instance, only one mean temperature estimate is given for England for the entire period AD 800–1000)³¹.

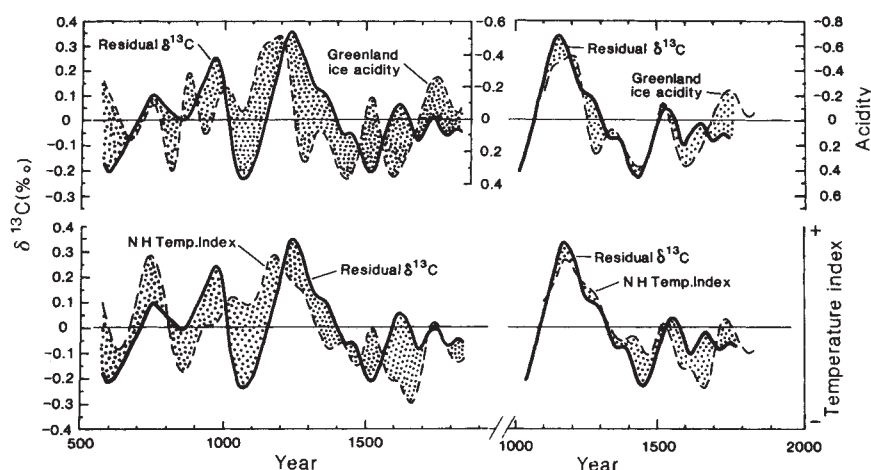


Fig. 4 Northern Hemispheric temperature index and Greenland ice acidity¹ (in deviations from the mean in μ equivalent $H^+ kg^{-1}$ ice) compared to the residual $\delta^{13}C$. The curves are without a lag on the left side and with a 70 to 90 yr shift for $\delta^{13}C$ on the right side (see text).

A cubic spline of Greenland ice acidity¹, which is a measure of volcanic activity and potential climatic change, is also given in Fig. 4, together with the tree cellulose δ_i curve. The NH index, δ_i values, and acidity units are all plotted such that increased y-axis values indicate increased temperature or reduced relative humidity. The curves were normalized on their variance, that is, the standard deviations around the mean are of equal length in the graphs.

An appreciable discrepancy between the NH temperature index and the δ_i curve is found near AD 1100 where the δ_i minimum is more pronounced. In addition, the δ_i derived curve is lacking the deep Little Ice Age minimum in the second half of the seventeenth century.

Significant correlations between δ_i and either the acidity or NH index ($r=0.00$ and 0.30 , respectively) are absent. Yet the AD 1100–1850 δ_i curve and the NH temperature index look very similar. Both curves have a maximum in the early part of the current millennium, as well as a parallel trend from about AD 1150 to 1500. The δ_i values appear to lag the NH temperature index by about 70 y. Moving the δ_i record 70 y back in time increases the 0.47 correlation coefficient of the post-AD 1100 interval to 0.79. Similarly, a 90 yr shift increases the δ_i -acidity correlation coefficient from 0.06 to 0.81 for the post-AD 1100 interval (Fig. 4, right).

Previously²⁶ we attributed the observed δ_i changes to the transfer of biospheric carbon to the atmosphere. Our δ_i -latitude relationship (Fig. 1) leads us to climate as another major variable influencing the pre-anthropogenic δ_i record. A combination of both factors is possible, with climate-induced biospheric reservoir size changes resulting in a perturbation of atmospheric (and tree-cellulose) $\delta^{13}C$. For such a mechanism δ_i would lag climate. Another possible explanation for the lag could be the slow response of mature trees to changing environmental conditions. The time lag of 70 to 90 y would then be caused by physiological adaptation to long-term climatic change.

Interpreting the latitudinal dependence of δ_i as humidity controlled leads to maximum pre-1850 relative humidity changes of $0.6/0.06 = 10\%$ in the current millennium for western North America.

The above palaeoclimatic considerations can be applied to glacial-interglacial wood $\delta^{13}C$. Yapp and Epstein³² investigated $\delta^{13}C$ of North American ^{14}C -dated wood ranging in age from 9,500 to 22,000 yr BP. The samples were subdivided into an interglacial (9,500–10,050 yr BP), transitional (10,200–12,920 yr BP) and glacial (14,300–22,400 yr BP) group. We calculate a latitudinal gradient of $-0.10 \pm 0.04\%$ per degree latitude ($r=0.43$, $n=38$, $p=0.01$) for these samples. Relative to our modern all-wood latitudinal δ_i gradient, the glacial and transitional group have, on average, lower δ_i values (0.8 and 1.8%, respectively). These lower δ_i values could have resulted from higher

relative humidities or lower temperatures. The interglacial group, however, has the largest average decline (2.4%). One would not have expected a δ_i lowering for this group. But few of these samples were identified with regard to species and it is possible that the anomaly is produced by a shift away from coniferous towards deciduous trees which are, at least in Fig. 1, 3.0% lower in δ_i .

The above discussion attempts to evaluate the factors controlling long-term $\delta^{13}C$ change in trees. A latitudinal relationship appears beyond dispute, and the record of δ_i change of the current millennium has patterns compatible with existing climate information. For glacial-interglacial wood samples, however, we have to agree with Yapp and Epstein³² that the palaeoclimatic significance of $\delta^{13}C$ is not yet fully resolved.

The carbon isotope research of the Quaternary Isotope Laboratory was supported by the U.S. Department of Energy.

Received 29 December 1986; accepted 16 April 1987.

1. Hammer, C. U., Clausen, H. B. & Dansgaard, W. *Nature* **288**, 230–235 (1980).
2. Francey, R. J. & Farquhar, G. D. *Nature* **297**, 28–31 (1982).
3. Freyer, H. D. *Proc. Int. Meet. Stable Isot. Tree-Ring Res.* New Paltz, New York, May 22–25, 1979 (ed. Jacoby, G. C.) 13–21 (US Dept of Energy, Washington DC, 1980).
4. Long, A. in *Climate from Tree Rings* (eds Hughes, M. K., Kelley, P. M., Pilcher, J. R. & LaMarche, V. C.) 12–18 (Cambridge University Press, 1982).
5. Stuiver, M. & Burk, R. L. in *Paleoclimate Analysis and Modeling* (ed. Hecht, A. D.) 151–161 (Wiley, New York, 1985).
6. Pearman, G. I., Francey, R. J. & Fraser, P. J. B. *Nature* **260**, 771–773 (1976).
7. Wilson, A. T. & Grinsted, M. J. *Nature* **265**, 133–135 (1977).
8. Freyer, H. D. & Belacy, N. *J. geophys. Res.* **88**, 6844–6852 (1983).
9. Tans, P. P. thesis, Univ. Groningen, Groningen, the Netherlands (1978).
10. Leavitt, S. W. & Long, A. *Tellus* **35B**, 92–102 (1983).
11. Leavitt, S. W. & Long, A. *Isotope Geoscience* **1**, 169–180 (1983).
12. Winter, K., Holtum, J. A. M., Edwards, G. E. & O'Leary, M. H. *J. exp. Botany* **33**, 88–91 (1982).
13. Winter, K. *Z. Pflanzenphysiol.* **101**, 421–430 (1981).
14. Francey, R. J. *Nature* **290**, 232–235 (1981).
15. Harkness, D. D. & Miller, B. F. *Radiocarbon* **22**, 291–298 (1980).
16. Lerman, J. C. & Long, A. *Proc. Int. Meet. Stable Isot. Tree-Ring Res.* New Paltz, New York, May 22–25, 1979 (ed. Jacoby, G. C.) 22–32 (US Dept of Energy, Washington DC, 1980).
17. Rebello, A. & Wagener, K. *Environmental Biogeochemistry* Vol. 1 (ed. Nriagu, J. O.) 13–24 (Ann Arbor Science, Ann Arbor, Michigan, 1976).
18. Tans, P. P. & Mook, W. G. *Tellus* **32**, 268–283 (1980).
19. Bender, M. & Berge, A. J. *Tellus* **34**, 500–504 (1982).
20. Grinsted, M. J., Wilson, A. T. & Ferguson, C. W. *Earth planet. Sci. Lett.* **42**, 251–253 (1979).
21. Galimov, E. M. *Environmental Biogeochemistry* Vol. 1 (ed. Nriagu, J. O.) 3–12 (Ann Arbor Science, Ann Arbor, Michigan, 1976).
22. Farmer, J. G. & Baxter, M. S. *Nature* **247**, 273–275 (1974).
23. Farmer, J. G. *Nature* **279**, 229–231 (1979).
24. Lerman, J. C., Mook, W. G. & Vogel, J. C. in *Radiocarbon Variations and Absolute Chronology* Nobel Symp. 12 (ed. Olsson, I. U.) 275–299 (Wiley, New York, 1970).
25. Hakansson, S. *Quat. Sci. Rev.* **4**, 135–146 (1985).
26. Stuiver, M., Burk, R. L. & Quay, P. D. *J. geophys. Res.* **89**, 11731–11748 (1984).
27. Stuiver, M. & Quay, P. D. *Science* **207**, 11–19 (1980).
28. Ramesh, R., Bhattacharya, S. K. & Gopalan, K. *Earth planet. Sci. Lett.* **79**, 66–74 (1986).
29. Freyer, H. D. *Proc. 6th Oak Ridge Life Sciences Symp.* Oak Ridge (eds Trabalka, J. R. & Reichle, D. E.) 125–151 (Springer, New York, 1986).
30. Gedzelman, S. D. *The Science and Wonders of the Atmosphere* 200–203 (Wiley, New York, 1980).
31. Dansgaard, W. in *Climatic Variations and Variability: Facts and Theories* (ed. Berger, A.) 193–206 (NATO Advanced Study Institute Series, Reidel, Dordrecht, 1980).
32. Yapp, C. J. & Epstein, S. *Earth planet. Sci. Lett.* **34**, 333–350 (1977).

A new fossil ctenophore discovered by X-rays

George D. Stanley Jr* & Wilhelm Stürmer†‡

* Department of Geology, University of Montana, Missoula, Montana 59812, USA

† Burgbergstrasse 20, D-8520 Erlangen, FRG

Ctenophores or comb jellies are tiny almost totally soft-bodied, bioluminescent planktonic organisms whose principal means of propulsion is the beating of eight ciliated bands which traverse their gelatinous exterior. Traditionally their distant ancestry has been thought to be linked to that of the Cnidaria, although recent detailed anatomical and ecological studies have been rallied to show that ctenophores have a history quite independent of cnidarians¹. Until recently, ctenophores were unknown as fossils and quite naturally were not expected to be found. However, we recently reported a remarkable cydippid-like specimen, *Paleoctenophora brasseli*², discovered in the Lower Devonian Hunsrück Slate of West Germany. This formation is now celebrated for its spectacularly well-preserved fossils, including the occasional preservation of tissue, organs and soft-parts³⁻⁵. Designated as *Archeocydippida hunsrueckiana* n. gen. n. sp., here we report an additional specimen discovered by X-rays. It reveals even more detail than our first specimen, including preservation of the delicate comb rows and possible gonads.

The first fossil ctenophore from the Devonian bore striking resemblance to extant forms and as the sole representative of an entire soft-bodied living phylum, was probably the rarest fossil. The presence of possibly eight meridional comb rows, aboral statocysts, and paired retractable tentacles suggested that *Paleoctenophora brasseli* was closely comparable to extant cydippids such as *Pleurobrachia*. Other problematic fossils have previously been suggested as possible ctenophores⁶⁻⁹ but closer study shows that this is unwarranted, owing to lack of correspondence to ctenophoran anatomy and greater resemblance to other groups.

Because ctenophores possess certain unique characteristics of more structurally advanced body organization, setting them

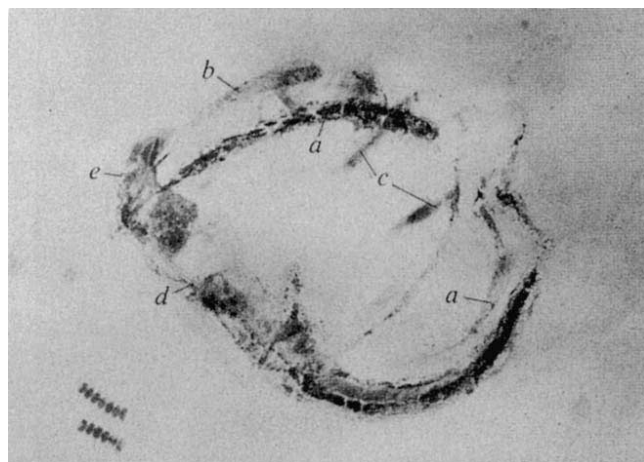


Fig. 2 *Archeocydippida hunsrueckiana* n. gen., n. sp., enlarged X-ray print of specimen in Fig. 1, $\times 4$, showing labelled anatomical details including a, the individual comb rows, some of which bear comb plates, several faintly visible on the left and possibly on the right; b, part of an individual comb row and possibly gonads (See Fig. 3); c, two possible tentacular bulbs; and d, a possible mouth. e, Sheet-like material adhering to oral surface may be decomposed muscle fibres or some type of extraneous material. The two segmented objects in the lower left of photo are unrelated crinoid ossicles.

apart from cnidarians, many workers consider them as an early, separate offshoot from a diploblastic stock. The presence of a fundamentally coherent ctenophoran body plan in the Devonian over 400 Myr ago, established the independent nature of the phylum, suggesting that their roots could extend even further back in time.

Shortly after publication of our first paper, a problematic soft-bodied fossil from the middle Cambrian Burgess Shale referred to as *Fasciculus*, was illustrated and suggested as a second fossil ctenophore¹⁰. A similar but larger fossil, also from the Burgess Shale, earlier had been assigned as *Fasciculus vesanus*¹¹ and was considered a problematic cnidarian. Like other Burgess Shale taxa, *Fasciculus* does not appear to be

‡ Deceased.

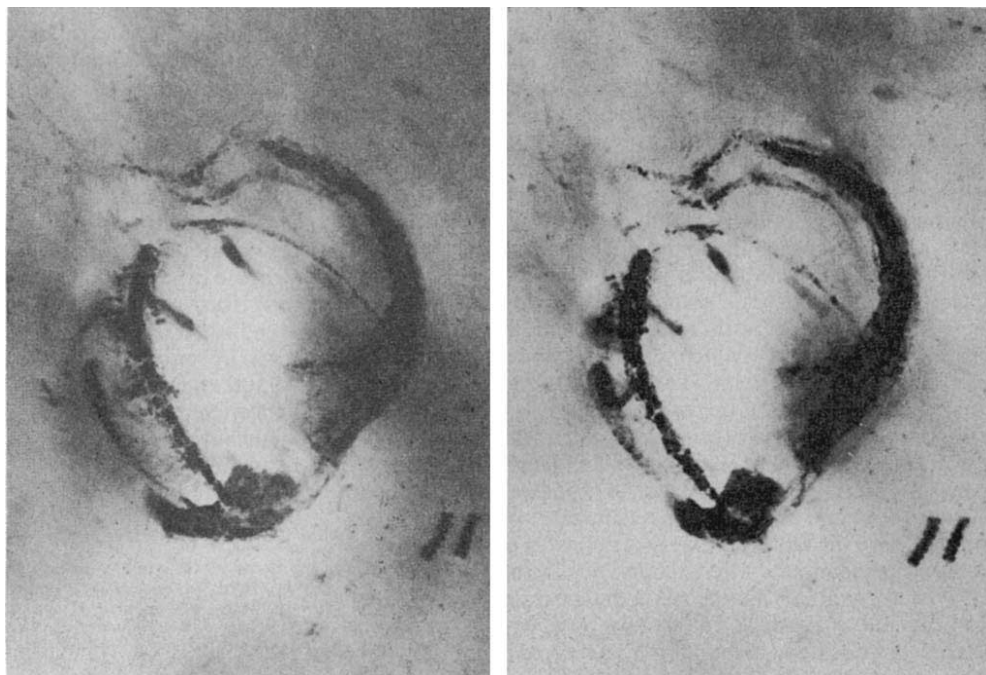


Fig. 1 *Archeocydippida hunsrueckiana* n. gen., n. sp. X-ray, stereo pair of specimen entombed in slate, $\times 2.5$. From the Hunsrück Slate, Kaisergrube quarry, Gemunden, West Germany. Repository: Bavarian State Museum collection, Munich, 1986 I3.



Fig. 3 Enlargement ($\times 12$) of distal portion of a comb row on right side of specimen showing spherical objects, possibly relating to gonads known on extant forms.

related to any living group. Its well-preserved anatomy is not at all compatible with *Paleoctenophora*¹². Although we have not examined the specimens, from the excellent photographs presented, we doubt, on the basis of currently accepted anatomical criteria whether these Middle Cambrian fossils can be considered as ctenophores. Aside from a globular shape and possible biradial symmetry, they appear to lack basic characteristics necessary for assignment to the Ctenophora^{1,13-15}. These include the possession of statocysts and a complex of eight meridional comb rows and fused macrociliated comb plates. Like a medusoid cnidarian, the Cambrian specimens reveal a plethora of tentacles, unlike the typical two tentacles and tentacular sheaths exhibited by our Devonian ctenophores and most extant species. Because many taxa of the Cambrian were in evolutionary stages of early experimentation¹⁶, they are frequently difficult to place in currently accepted phyla or classes. But we prefer to regard the multi-tentaculated *Fasciculus* as a problematic cnidarian, defying classification with extant groups. We concede that one could postulate that these Cambrian specimens are ancestral stock leading to true ctenophores, but without more data, especially from the time interval between early Devonian and middle Cambrian, this assertion remains to be proved.

As revealed by X-rays, we illustrate here a new ctenophore from the Hunsrück Slate. It is 23 mm long and compared to *Paleoctenophora brasseli*, shows a greater degree of anatomical disarray, perhaps associated with flattening and partial decomposition. A possible mouth, two obviously paired tentacular sheaths and possibly eight clearly differentiated row combs, several of which still bear the delicate, ciliated comb plates (Figs 1 and 2) are discernible in the specimen. Although over-pyritization of the other comb plates has obliterated some of their details, such resolution was not possible with the first Hunsrück ctenophore. The distal (oral) portion of one comb row bears what appears to be clumps and linear arrays of spherical objects (Fig. 3) which may be individual comb plates but which also bear striking similarity to the gonads of some extant forms¹⁷. We assign this second specimen to *Archeocydippida hunsrueckiana*. With the shape and two tentacular bulbs clearly discernible, our specimen may be placed confidently in the order Cydippida, class Tentaculata. The tentacles appear to exit near the aboral pole in distinct contrast to *Paleoctenophora brasseli*, where the tentacles exit through short tentacle sheaths located near the mouth. The anatomy discernible in our specimen even appears sufficient to place the specimen in the family Pleurobrachiidae although if it is compressed in the stomodaeal axis and the position of the mouth is correctly interpreted, it may belong to the Mertensiidae, another extant family.

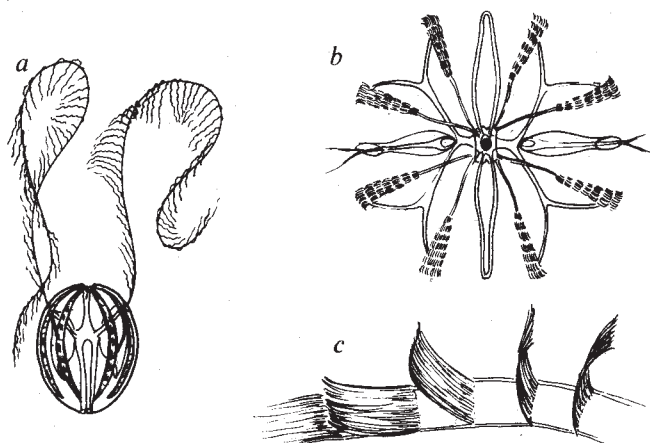


Fig. 4 Sketch of the basic anatomy of a living cydippid ctenophore *Pleurobrachia*, so abundant and widespread in temperate waters of the North Atlantic. a, Whole organism with tentacles and tentacular sheath; b, aboral view showing eight rows of overlapping comb plates; and c, detailed sketch of a portion of a row, showing the ciliated plates which beat in wave-like motions to propel the organism. (Sketches after Hyman¹³.)

Because many other Hunsrück fossils like our ctenophore are preserved in pyrite, they can be investigated using X-radiography. The presence of the flattened fossil ctenophore illustrated here was first detected with X-rays. Once the exact position of the entombed specimen was ascertained, the slab of black slate in which it is entombed was ground to a thickness of about 2 mm and X-rayed with a modified Siemens Kristalloflex II water-cooled, fine focus X-ray unit. The photographs were produced with Agfa Millimask plates (similar to Kodak maximum resolution plates). Typical exposure was for ten hours at 40 kV, 10 mA. The reason for such a long exposure period is the extremely thin emulsion with a very low absorption for X-ray quanta. The glass plates produced were inspected with a high-resolution TV microscope in combination with a modified Siemens Transicon for variable enhancement of details. As the fossils of the Hunsrück Slate are invariably flattened, the angle for stereo-radiographs was increased to $\pm 15^\circ$. With such exposures the flattened object is 'decompressed', providing a better idea of the original three-dimensional shape (Fig. 1). Without the sophisticated X-ray techniques used, both of these rare specimens would not have been detected.

Comparing closely to extant forms, our new fossil ctenophore constitutes the second fossil example. As revealed by X rays, it is unique in providing additional details of the comb plates and possible gonads (Figs 2 and 3). The two Devonian specimens, over 400 Myr old, show little anatomical deviation from extant families and together they reaffirm the distinction of the basic ctenophoran body organization over that of the Cnidaria. Because these ancient ctenophores are little differentiated from living forms, it suggests that the origin of the phylum may extend even further back in time. The clearly cydippid anatomy of both these Devonian fossils is regarded as primitive and ancestral to all subsequent stock alive today^{1,13}. Although only two fossil taxa are now known, like today, seas of the early Devonian time might have been highly populated by a variety of these gelatinous planktonic organisms.

Received 22 January; accepted 16 April 1987.

1. Harbison, G. R. in *The Origins and Relationships of Lower Invertebrates* (eds Conway Morris, S., George, J. D., Gibson, R. & Platt, H. M.) 78-100 (Systematics Ass. Spec. Vol. 28, London, 1985).
2. Stanley, G. D. Jr & Stürmer, W. *Nature* **303**, 518-520 (1983).
3. Stürmer, W. *Science* **170**, 1300-1302 (1970).
4. Yochelson, E. L., Stürmer, W. & Stanley, G. D. Jr *Paläont. Z.* **57**, 39-68 (1983).
5. Stürmer, W. *Nature* **318**, 53-55 (1985).

6. Moore, R. C. in *Treatise on Invertebrate Paleontology* Part O (ed. Moore, R. C.) O16-O21 (Geological Society of America and University of Kansas Press, Lawrence, Kansas, 1959).
7. Caster, K. E. & Brooks, H. K. *Bull. Am. Paleont.* **36**, 157-199 (1956).
8. Häntzschel, W. in *Treatise on Invertebrate Paleontology* Part W (ed. Teichert, C.) W1-W269 (Geological Society of America and University of Kansas Press, Lawrence, Kansas, 1975).
9. Smith, A. B. *Alcheringa* **6**, 93-99 (1982).
10. Collins, D., Briggs, D. & Conway-Morris, S. *Science* **222**, 163-167 (1983).
11. Simonetta, A. M. & Delle Cave, L. *Atti Soc. Toscana Sci. Nat. Pisa Mem. Ser.* **85**, 45-49 (1978).
12. Conway Morris, S. in *The Origins and Relationships of Lower Invertebrates* (eds Conway Morris, S., George, J. D., Gibson, R. & Platt, H. M.) 343-359 (Systematics Ass. Spec. Vol. 28, London, 1985).
13. Hyman, L. H. *Invertebrates: Protozoa through Ctenophora* (McGraw-Hill, New York, 1940).
14. Harbison, G. R., Madin, L. P. & Swanbert, N. R. *Deep Sea Res.* **25**, 233-255 (1978).
15. Barrington, E. J. W. *Invertebrate Structure and Function* (Nelson, Sunbury-on-Thames, 1979).
16. Whittington, H. B. *The Burgess Shale* (Yale University Press, New Haven, 1985).
17. Harbison, G. R. & Miller, R. L. *Mar. Biol.* **90**, 413-424 (1986).

Elucidation of the two-dimensional structure of an α -amino acid surfactant monolayer on water using synchrotron X-ray diffraction

Sharon Grayer Wolf*, Leslie Leiserowitz*, Meir Lahav*, Moshe Deutsch†, Kristian Kjaer‡ & Jens Als-Nielsen‡

* Structural Chemistry Department, Weizmann Institute of Science, Rehovot 76100, Israel

† Physics Department, Bar-Ilan University, Ramat Gan 52100, Israel

‡ Physics Department, Risø National Laboratory, DK4000 Roskilde, Denmark

Recently it has been demonstrated that a compressed monolayer of palmitoyl-(R)-lysine (Fig. 1) on the surface of water induces the orientated growth of α -glycine crystals¹. This effect was interpreted to be a structural match between the monolayer and the α -glycine. To test this hypothesis we undertook to determine the packing arrangement of the monolayer by grazing incidence X-ray diffraction and reflection. These techniques² use the unique properties of synchrotron radiation: high intensity within a small natural collimation³. In the diffraction experiment two peaks were detectable in the two-dimensional powder pattern from a monolayer of palmitoyl-(R)-lysine. The positions and intensities of these peaks allowed us to choose between various models and determine the monolayer structure. This is the first time that the crystal structure of a compressed surfactant monolayer at the air-water interface has been determined. The same techniques could be used for structural characterization of other monolayers of interest in fields as diverse as biological membranes⁴ and optical second harmonic generation⁵. The packing arrangement of the α -amino acid head groups in the model proved to be very similar to that found in the crystal structures of α -glycine and several hydrophobic α -amino acids.

Crystalline α -glycine has a packing arrangement⁶ consisting of hydrogen-bonded centrosymmetric bilayers. The orientated binding of α -glycine crystals under the resolved surfactant monolayer has been explained by the formation of an analogous bilayer between the surfactant monolayer and an underlying layer of glycine molecules via a pseudo centre of symmetry as depicted in Fig. 1B. The determination of a structure of the monolayer that is similar to that of the α -glycine would verify this theory. Surface X-ray diffraction techniques using synchrotron radiation have been used to study the overlayer structure on single crystals^{7,8} as well as reconstructed surfaces^{9,10}. Studies of the surface of water¹¹ indicated that it would be possible to obtain in-plane diffraction from a monolayer at the air-water interface. Indeed, a diffraction peak has been detected for a phospholipid monolayer¹². Here, in parallel studies, we have obtained two peaks from in-plane diffraction of a compressed monolayer of palmitoyl-(R)-lysine on the surface of water. The additional information provided by the second peak has allowed, for the first time, a distinction to be made between

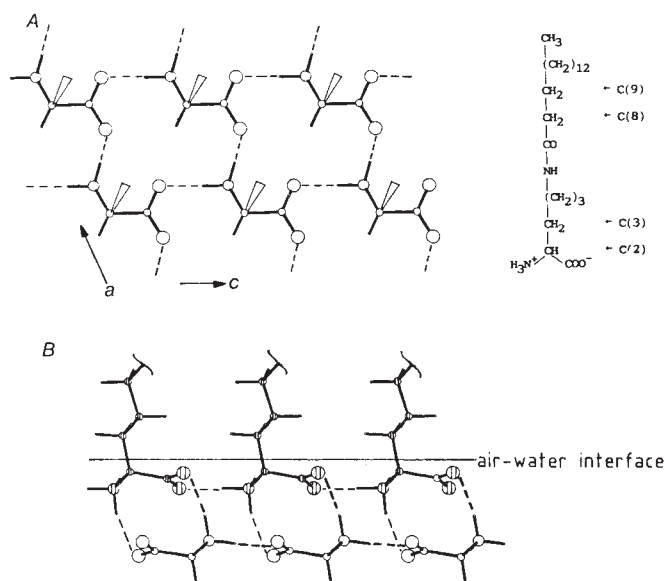


Fig. 1 A, The packing arrangement of the surfactant palmitoyl-(R)-lysine monolayer as derived from the packing arrangement of α -glycine⁶ in a view projected onto the monolayer plane. Hydrogen-bonding within the layer is represented by dashes. The wedge represents the projected side chain. The molecule palmitoyl lysine is shown on the right. B, Side view of the surfactant-glycine bilayer. The molecules with darkened atoms represent the surfactant monolayer. The molecules with light atoms represent the monolayer of α -glycine.

different possible models for the monolayer packing arrangement and thereby selection of the one best fitting the data.

Palmitoyl lysine was designed to be a surfactant ideal for the study of surface structure. Its glycine moieties ($^+H_3N-CH-COO^-$) are interlinked in the monolayer along the a - and c -axes (Fig. 1) by $NH\cdots O$ hydrogen bonds. The secondary amide group (CONH) incorporated in the chain of the molecules, contributes to the structural stability by interlocking neighbouring molecules via $NH\cdots O=C$ bonds (see Fig. 3B). These hydrogen bonds have a translation repeat compatible with the lattice repeat distances of the glycine moiety. The N-atom of the secondary amide group is placed an even number (four) of carbon atoms away from the amino-acid group so that the $NH\cdots O$ hydrogen bonding is optimized. Then the N-H points towards the oxygen of a neighbouring molecule, whereas an odd number of carbon atoms between the two groups would result in a nonlinear intermolecular $NH\cdots O$ contact of about 4 Å, precluding hydrogen bonding. The backbone bonds in the system $CH_2CH_2NHC=OCH_2CH_2$ can be taken to be generally all *trans*, according to the structures of peptides and aliphatic amides¹³, and model compounds of nylon¹⁴.

Grazing incidence X-ray diffraction and specular reflectivity measurements were performed on compressed monolayers of palmitoyl-(R)-lysine at the air-liquid interface using the liquid surface diffractometer at beamline D4, Hasylab, Hamburg. The reflectivity technique has been described previously¹⁵. The grazing incidence diffraction was performed on the same experimental set-up as found in ref. 12. We used a wavelength of 0.138 nm and a grazing angle of 0.85 times the critical angle for total reflection (0.137°) yielding a penetration depth of 8.7 nm of the evanescent wave. The in-plane diffraction results are shown in Fig. 2. Sub-phases of undersaturated solutions of glycine and (S)-glutamine were found to stabilize further the monolayer, revealed by a slowing of surface tension decay. The stabilization can occur possibly via bilayer formation at the air-solution interface, as shown in Fig. 1. There was no obvious difference in the diffraction patterns obtained from the

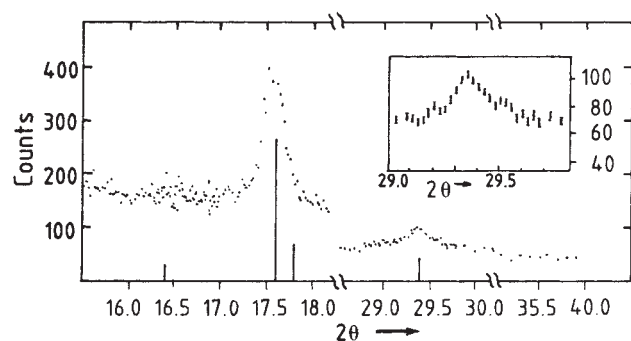


Fig. 2 Experimental X-ray diffraction pattern from a multi-domain monolayer of palmitoyl-(R)-lysine. The calculated relative intensities for conformer (i) are drawn in underneath the peaks. The inset shows an expanded view of the (1, 2) peak. The angle of incidence of the exploring beam was fixed at a value slightly below θ_c , the critical angle for total external reflection, and the intensity scattered horizontally parallel to the plane of the monolayer scanned. The measured horizontal full width at half maximum resolution of $\Delta(2\theta) = 2.6$ mrad was largely determined by the Soller collimator employed in front of the detector. The samples were prepared in a Langmuir trough equipped with a Wilhelmy balance which was contained within a gas-tight canister with X-ray transparent windows. The trough was equipped with a water thermostat, and was kept at a temperature of 19 °C while the ambient room temperature was 23–25 °C.

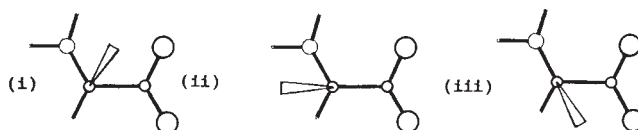
monolayer over the subphase from that over pure water. Reflectivities of >96% were obtained for water and for the monolayer on water, indicating exceptionally smooth surfaces¹¹.

In-plane diffraction revealed peaks at 2θ values of 17.55° and 29.4°, corresponding to d -spacings of 4.5 and 2.7 Å respectively, with an intensity ratio of $8 \pm 1:1$ (see Fig. 2). The 2.7-Å spacing rules out hexagonal close packing for the monolayer since $4.5/2.7 \neq \sqrt{3}$. A mosaic scan, where the detector angle (2θ) is set to the first peak position and the sample rotated through 60°, revealed no significant intensity changes, showing the sample to be a two dimensional powder. Using the Scherrer formula¹⁶ and taking into account the measured in-plane resolution, the width of either peak yields a coherence length of 500 ± 100 Å. The peak around $2\theta = 17.5^\circ$ was measured at surface pressures from 0 to 50 mN m⁻¹ in five different experiments. The peak first appeared at between 15 and 25 mN m⁻¹ corresponding to a molecular surface area ranging between 24 and

26 Å² per molecule and thereafter the position did not display any pressure dependence.

Structural chemical knowledge can be used to derive a limited number of possible models for the two-dimensional structure of the palmitoyl-(R)-lysine monolayer. Single crystal studies of valine¹⁷, leucine¹⁸, isoleucine¹⁹, norleucine²⁰ and (S)- α -amino-N-octanoic acid (measured in our laboratory) reveal that chiral resolved hydrophobic amino acids pack in layers in which only translation symmetry is used. The layer arrangement of α -glycine⁶ (see the layer in Fig. 1) is the prototype for the hydrophobic amino acids, with cell axes of $a = 5.10$, $c = 5.46$ Å, and $\beta = 111.7^\circ$, and a corresponding layer area of 25.9 Å² per molecule. Atom-atom potential energy calculations have shown that the α -glycine monolayer arrangement is the most stable (has the lowest energy) of the possible layers constructed by translation symmetry²¹. A modification of this arrangement was used as a prototype for the palmitoyl-(R)-lysine monolayer. The electrostatic parameters of the glycine and amide groups used to calculate the Coulomb component of the energy were obtained from deformation electron densities of α -glycine²² and amide crystal structures²³, derived from low-temperature X-ray diffraction studies. The deformation density is the electron density of the bonded atom minus that of the free atom. Parameters of this type have been used successfully for the analysis of packing arrangements and crystal morphology of amides and carboxylic acids^{24–26}.

A model for the molecular geometry of palmitoyl-(R)-lysine was constructed from the crystal structure of (S)- α -amino octanoic acid (⁺H₃N-CH(C₆H₁₃)-COO⁻) whose linear chain C₆H₁₃ is all *trans* and corresponds to conformer (i) below. The molecule appears only in three conformational regions with respect to rotations around the C(2)–C(3) bond, each 120° apart and all in staggered conformations (i), (ii) and (iii) below.



These three conformations have been observed in the structures of (S)- α -amino octanoic acid, (S)-norleucine, and (R, S)- α -amino-N-butyric acid²⁷, respectively. These three possible conformers are compatible with the monolayer packing arrangement in that they correspond to the three possibilities for the secondary amide groups in the chains to hydrogen-bond along $a + c$,

Table 1 The strongest reflections (calculated and observed)

h	l	2θ	$d(\text{\AA})$	$I(i)$	Calculated*	$I(iii)$	Observed	$d(\text{\AA})$
0	1	16.4	4.84	10 (112, 12)	3 (2, 2)	888 (1797, 212)	<10	—
1	$\bar{1}$	17.6	4.50	79 (30, 113)	0.2 (0.2, 0.2)	56 (54, 65)	100	4.53
1	0	17.8	4.45	21 (15, 28)	99.8 (44, 97)	44 (72, 87)		
1	$\bar{2}$	29.4	2.72	12 (9, 27)	0.4 (0.3, 0.2)	12 (9, 9)	12 $\pm$ 1.5	2.72

A tabulation of peak positions and intensities for the observed monolayer and the three calculated models (i), (ii) and (iii). Intensities are scaled such that $I[(1, 0) + (1, \bar{1})] = 100$ for these conformers. Numbers in brackets are intensities calculated for the particular reflection after rotating the molecule around the C2–C3 bond +5° and –5° respectively. Positive rotation is clockwise when looking from C2 to C3. Here is the formula used to calculate the intensities of the packing arrangements of conformers (i), (ii) and (iii).

$$* I_{hl} = |F_{hl}|^2 \cdot \frac{1}{\sin 2\theta} \cdot \frac{\cos^2 2\theta}{\sin 2\theta} \text{ where } \frac{1}{\sin 2\theta} = \text{beam cross-section, area factor } \frac{\cos^2 2\theta}{\sin 2\theta} = L_p \text{ factor, and } F_{hl} \text{ is the structure factor given by; } F_{hl} = \sum f_j \left(\frac{\sin \theta}{\lambda} \right) \exp[-(8\pi^2 U \sin^2 \theta / \lambda^2) \exp[2\pi i(hx_j + lz_j)]]$$

Fig. 3 *A*, Stereoscopic view perpendicular to the monolayer plane showing the packing arrangement for the palmitoyl-(R)-lysine monolayer. *B*, Stereoscopic view along the axis a - c showing molecules in a row parallel to the $a+c$ direction, interlinked by $N-H\cdots O=C$ (amide) hydrogen bonds. The secondary amide hydrogen-bonding is denoted by the dashes. The 24.8 Å spacing represents the distance between the methyl carbon and the amino nitrogen. The 27.4 Å spacing includes van der Waals radii totaling 2.6 Å.

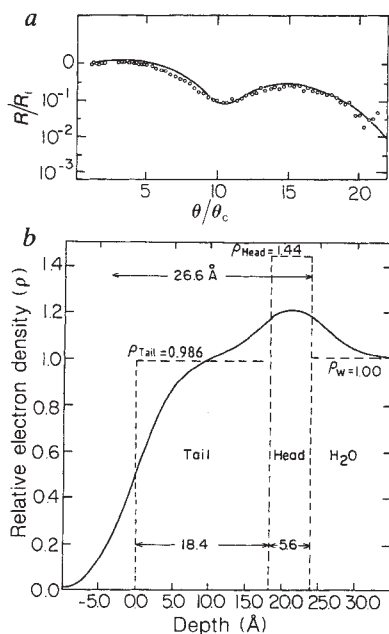
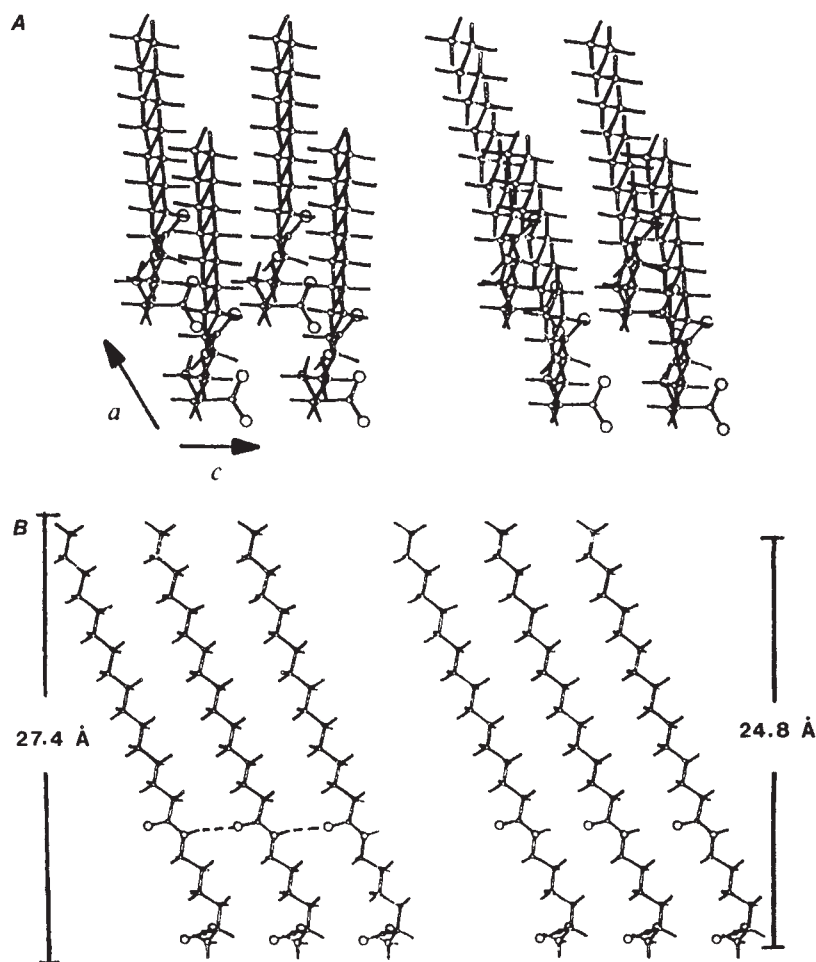


Fig. 4 *a*, Reflectivity data (···) and fitted model (—) for palmitoyl-(R)-lysine. R , measured reflectivity; R_i = Fresnel reflectivity of an ideally smooth and abrupt surface². *b*, Simplified model of the monolayer electron density (ρ) relative to that of water. The profile (full line) results from convoluting a box model of the density (broken lines) with a gaussian of standard deviation $\sigma = 4.4$ Å. The fitted layer thickness of 26.6 Å is indicated.

c or a . The first possibility is the one shown in Fig. 3. Other conformations are ruled out by both intramolecular and intermolecular energy considerations.

Palmitoyl-(R)-lysine molecules were then fitted into the ac layer structure of α -glycine and the cell constants a , c , and the angle β were varied to yield the lowest monolayer lattice energy (-414 kJ mol⁻¹). The model layer arrangement is shown in Fig. 3, with cell constants of $a = 5.03$ $c = 5.46$ Å and $\beta = 117.8^\circ$, and a corresponding molecular area of 24.3 Å² per molecule. This is substantially smaller than the area of the α -glycine layer (25.9 Å² per molecule) but is to be expected since hydrocarbon chains alone pack with an area of 20 Å² per molecule. Indeed, the molecular area of the shorter chained (S)- α -amino octanoic acid (25.0 Å² per molecule) lies between the above two values. The C2—C3 bond is tilted 11.4° from the normal to the monolayer surface in the direction parallel to and away from the C(carbonyl)—C2 bond. The tilt of the bond in this direction is found in the hydrophobic amino acid crystal structures of (R,S)- and (S)-leucine, (S)-isoleucine, (R,S)- and (S)-valine and (S)- α -amino octanoic acid with angles ranging from 10 to 17°. In α -glycine the C2—H bond has a tilt angle of 10.4° in the same direction as stated above. Conformer (i) in the palmitoyl lysine arrangement causes the hydrogen-bonded amide group in the side chains to lie almost parallel to the $a+c$ axis, whose length of 5.4 Å allows a long but not unreasonable $NH\cdots O$ hydrogen bond of 3.27 Å²⁸. Calculation shows this hydrogen bond to be only about 1.2 kJ mol⁻¹ less stable than is the normal bond with a $NH\cdots O$ distance of 2.95 Å²⁵. The monolayer energies of conformers (ii) and (iii) were at least 25–30 kJ mol⁻¹ higher (less stable) than that of conformer (i).

These energy differences alone are not large enough for an unambiguous choice between the conformers. However, the

calculated two-dimensional diffraction patterns of the various conformers are very different. Table 1 compares the models to the experimental data (Fig. 2). The data eliminate conformers (ii) and (iii), in which the molecular arrangements are such that the amide groups are interlinked by hydrogen bonds along the *a*- or *c*-axis. Even rotating the chain about the C2—C3 bond by $\pm 5^\circ$ does not improve the fit for these two models (see the calculated intensities in brackets). Their calculated intensities $I(h, l)$ using the model atomic (*x, z*) coordinates (see Table 1) do not match the experimental results. In particular, conformer (iii) would be associated with a very large (0, 1) peak, contrary to observation. Conformer (i) gives a good match to the observed intensities, in particular after the final section of the chain was rotated around the C8—C9 bond by 7.5° from the all *trans* conformation. The (0, 1) reflection is not expected to be detectable because of higher background intensity at low 2θ values. The 8:1 intensity ratio of the two observed peaks indicates a value of $U = 0.25 \text{ \AA}^2$ for the average isotropic thermal parameter (for all atoms). This temperature factor seems reasonable considering that the average isotropic thermal parameter is $0.14 \text{ \AA}^2$ for crystalline deoxylysophosphatidylcholine monohydrate²⁹.

The reflectivity data (Fig. 4a) was analysed by fitting to it the calculated reflectivity² of a simple model of the electron density variation across the monolayer (Fig. 4b). The resulting layer thickness of $26.6 \pm 1.0 \text{ \AA}$ is shown in Fig. 4b and compares well with the layer thickness of $27.2 \pm 0.2 \text{ \AA}$ provided by X-ray powder diffraction data of crystalline palmitoyl-(R)-lysine (measured by P. Spieker of Rigaku Co.). This, in turn, matches the layer thickness calculated for the model molecule ($27.4 \text{ \AA}$), as indicated in Fig. 3B. Note that the layer thicknesses quoted here include van der Waals radii totalling $2.6 \text{ \AA}$ and are not simply the distance between extremal non-hydrogen atom centres. The calculated model density of the monolayer is then 0.96 Mg m^{-3} compared with the measured density of crystalline palmitoyl-(R)-lysine of 1.03 Mg m^{-3} . If there were three-dimensional crystallites at the air-water interface then they would be platelets lying with the (010) face parallel to water. One would then detect the extremely intense (020) reflection ($d = 27.2 \text{ \AA}$) at $\theta/\theta_c = 10.6$ in the reflectivity curve. This was never observed. Thus the monolayer diffraction peaks observed must arise from two-dimensional structure.

The diffraction results indicate that the packing arrangement of the amino acid head groups in the monolayer is very close to that found in α -glycine and several hydrophobic α -amino acids. This explains the fact that the monolayer at the air-water interface induces epitaxial nucleation of these amino acids in solution. The ability to determine the size and structure of small domains will perhaps enable determination of the smallest domain size of the monolayer required to induce crystallization as well as the structure and size of the smallest nucleating crystal. Eventually we hope to monitor the process of nucleation of α -glycine onto the monolayer by real-time experiments involving the reflectivity and in-plane diffraction measurements. The possibility of inducing large monolayer domains (perhaps $4\text{--}5 \text{ mm}^2$) would allow for a less ambiguous 'single crystal' structure determination of the α -amino surfactant. It would also permit the direct assignment of its absolute configuration by simply noting the absolute angular sequence of the diffraction peaks, as was pointed out previously³⁰.

We thank Mourits Nielsen, Francois Grey, Ehud Landau, Ronit Popovitz-Biro and Karlheinz Hill for their help and advice; Peter Pershan, Jacob Sagiv, Mendel Cohen, Ziva Berkovitch-Yellin and Lia Addadi for stimulating discussions, Peter Spieker, Rigaku Corporation, for the X-ray powder diffraction spectra. We thank US/Israel Binational Science Foundation, Jerusalem, the fund for basic research of the Israel Academy of Sciences and Humanities, Stiftung Volkswagenwerk, the Petroleum Fund of the American Chemical Society, and the Danish Foundation for Natural Sciences for financial support.

Received 5 December 1986; accepted 2 April 1987.

- Landau, E. M., Levanon, M., Leiserowitz, L., Lahav, M. & Sagiv, J. *Nature* **318**, 353–356 (1985).
- Als-Nielsen, J. in *Structure and Dynamics of Surfaces* Vol. 2 Ch. 5 (eds Schommers, W. & von Blanckenhagen, P.) (Springer, Berlin, 1986).
- Buras, B. & Tazzari, S. *European Synchrotron Radiation Facility: Report of the ESRP 3rd edn* (CERN, Geneva, 1985).
- Weis, R. M. & McConnell, H. M. *Nature* **310**, 47–49 (1984).
- Berkovic, G., Rasing Th. & Shen, Y. R. *J. chem. Phys.* **85**, 7374–7376 (1986).
- Almlöf, J., Kvick, Å. & Thomas, J. O. *J. chem. Phys.* **59**, 3901–3906 (1973).
- Eisenberger, P. & Marra, W. C. *Phys. Rev. Lett.* **46**, 1081 (1981).
- Seul, M., Eisenberger, P. & McConnell, H. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5795–5797 (1983).
- Robinson, I. K. *Phys. Rev. Lett.* **50**, 1045 (1983).
- Bohr, J. *et al. Phys. Rev. Lett.* **54**, 1275–1278 (1985).
- Braslau, A. *et al. Phys. Rev. Lett.* **54**, 114–117 (1985).
- Kjaer, K., Als-Nielsen, J., Helm, C. A., Laxhuber, L. A. & H. Möhwald, H. *Phys. Rev. Lett.* (in the press).
- Dunitz, J. D. & Strickler, P. in *Structural Chemistry and Molecular Biology* (eds Rich, A. & Davidson, N.) (Freeman, San Francisco, 1968).
- Benedetti, E., Ciajolo, M. R. & Corradini, P. *Eur. Pol. J.* **9**, 101 (1973).
- Als-Nielsen, J. & Pershan, P. S. *Nucl. Instrum. Meth.* **208**, 545–548 (1983).
- Guinier, A. *X-ray Diffraction* Ch. 5.1 (Freeman, San Francisco, 1963).
- Torii, K. & Iitaka, Y. *Acta Crystallogr.* **B26**, 1317–1326 (1970).
- Torii, K. & Iitaka, Y. *Acta Crystallogr.* **B27**, 2237–2246 (1971).
- Harding, M. M. & Howieson, R. M. *Acta Crystallogr.* **B32**, 633–634 (1976).
- Torii, K. & Iitaka, Y. *Acta Crystallogr.* **B29**, 2799–2807 (1973).
- Wolf, S. G. thesis, Weizmann Institute of Science (1987).
- Legros, J. P. & Kvick, Å. *Acta Crystallogr.* **B36**, 3052–3059 (1980).
- Berkovitch-Yellin, Z., Leiserowitz, L. & Nader, F. *Acta Crystallogr.* **B33**, 3670–3677 (1977).
- Berkovitch-Yellin, Z. *J. Am. chem. Soc.* **107**, 8239–8253 (1985).
- Berkovitch-Yellin, Z. & Leiserowitz, L. *J. Am. chem. Soc.* **102**, 7677–7690 (1980).
- Berkovitch-Yellin, Z. & Leiserowitz, L. *J. Am. chem. Soc.* **104**, 4052–4064 (1982).
- Nakata, K., Takaki, Y. & Sakurai, K. *Acta Crystallogr.* **B36**, 504–506 (1980).
- Leiserowitz, L. *Acta Crystallogr.* **B32**, 775–802 (1976).
- Hauser, H., Pascher, I. & Sundell, S. *J. molec. Biol.* **137**, 249–264 (1980).
- Addadi, L., Berkovitch-Yellin, Z., Weissbuch, I., Lahav, M. & Leiserowitz, L. in *Topics in Stereochemistry* Vol. 16 (eds Eliel, E. L., Wilen, S. H. & Allinger, N.) (Wiley, New York, 1986).

Short-term selection for recombination among mutually antagonistic species

Graham Bell* & J. Maynard Smith†

* Biology Department, McGill University, Montreal, Quebec, H3A 1B1 Canada

† School of Biological Sciences, University of Sussex, Falmer, Brighton, East Sussex BN1 9QG, UK

The recombination of genes is clearly disadvantageous in a uniform and predictable environment where a single multilocus genotype is optimal. But heterogeneity or unpredictability do not necessarily imply a short-term advantage for recombination, which will be created only if the correlation between adaptively important features of the environment frequently changes sign¹. Another way of expressing this conclusion is to say that genes causing non-zero rates of recombination will tend to spread only if selection often favours a change in the sign of linkage (gametic phase) disequilibrium². Because it is difficult to imagine that the physical environment often behaves in this fashion, recent theoretical work on the maintenance of sex, recombination and outcrossing has speculated that biotic interactions may be important^{3–5}. May and Anderson⁶ analysed a two-locus haploid model of the coevolution of host and parasite, similar to that considered here, and showed that the mean fitness of a population with recombination is typically higher than that of a population without, although they did not explicitly consider changes in the frequency of genes affecting recombination. Hutson and Law⁷ and Bell⁸ suggested that the crucial feature of antagonistic interactions between species is the effect on the current fitness of a genotype of its frequency in previous generations; such time-lagged frequency-dependent selection tends to destabilize both genotype frequencies and linkage disequilibrium, whose oscillation has been shown to fuel the spread of high-recombination alleles in explicit genetic models⁸. In this paper, we shall show how the mutual antagonism of two species can lead to a cyclical game in which high-recombination alleles can have a large short-term selective advantage in a fully defined genetic model.

Because the parasite-host interaction is often highly specific and has major effects on fitness, we shall refer to the two antagonists as 'host' and 'parasite', although in principle our

		Response of parasite						Response of host			
Parasite genotype	AB	0.5	0.75	0.75	1		AB	1	0.75	0.75	0.5
	Ab	0.75	1	1	0.75		Ab	0.75	0.5	0.5	0.75
	aB	0.75	1	1	0.75		aB	0.75	0.5	0.5	0.75
	ab	1	0.75	0.75	0.5		ab	0.5	0.75	0.75	1
		ab	aB	Ab	AB			ab	aB	Ab	AB
Parasite genotype	AB	0.5	0.75	0.75	1		AB	1	0.75	0.75	0.5
	Ab	0.75	0.5	1	0.75		Ab	0.75	1	0.5	0.75
	aB	0.75	1	0.5	0.75		aB	0.75	0.5	1	0.75
	ab	1	0.75	0.75	0.5		ab	0.5	0.75	0.75	1
		ab	aB	Ab	AB			ab	aB	Ab	AB
		Host genotype									

Host genotype

Fig. 1 The response of host and parasite types when challenged by an antagonist. The two upper tables define a quantitative model in which *A*, *B* have unit genotypic value and *a*, *b* zero, with response depending on the absolute difference between the summed genotypic values of the two antagonists; consequently, *Ab* and *aB* have the same (unit) phenotype. The two lower tables define a 'gene for gene' model in which each gene in one species is specifically antagonized by a gene in the other; as a result *Ab* and *aB* have opposite phenotypes. Note that gene effects are additive in both models, which have been set up so as to give a twofold range in response; so long as the additive property is preserved, the direction of selection on the R^+ gene does not depend on the particular numerical values used.

model applies equally well to competitors or to predators and prey. Both parasite and host are sexual haplonts with a single chromosome bearing three loci arranged in the order R - A - B . The *A* and *B* loci determine viability or fecundity through the host-parasite interaction, whereas the only effect of alleles at the R locus is to alter the probability of crossing-over, with a dominant allele R^+ causing free recombination between the three loci and a recessive R^- completely suppressing recombination. The effects of changing these assumptions about the R locus are described below.

The fitness loci bear the pairs of alleles *A*, *a* and *B*, *b*, which determine the response of parasite to host and vice versa. This interaction can be modelled in two ways, depending on the properties of the repulsion genotypes *Ab* and *aB*. The first is to assume that resistance and virulence depend on the sum of effects across loci, with the alleles contributing larger or smaller effects to a total score. The crucial feature of this 'quantitative' model is that *Ab* and *aB* have identical properties. The second possibility is that genes at one locus in either partner are specifically antagonized by genes at a corresponding locus in its antagonist, leading to a 'gene-for-gene' mode of interaction'. The crucial feature of the gene-for-gene model is that *Ab* and *aB* have opposite properties. The matrices defining the response of the antagonists to one another in both models are shown in Fig. 1. To calculate overall fitness in either case, we assume that a host is infected by only one type of parasite, with probability proportional to the frequency of that type in the parasite population, and that the overall fitness of a given type of host is equal to the sum of its responses to all the parasite types, weighted by their frequencies. The fitness of parasite genotypes is calculated in the same way. We shall show results for both quantitative and gene-for-gene models in which gene effects are additive, using the parameter values given in Fig. 1.

These assumptions have been investigated by computer simulation, starting with a low frequency of the R^+ allele (10^{-5}) in both species, with the frequencies of the fitness alleles close to 0.5. Initial gamete frequencies had a small random departure from linkage equilibrium: replicate runs showed that the results are not altered qualitatively by this variation in initial conditions.

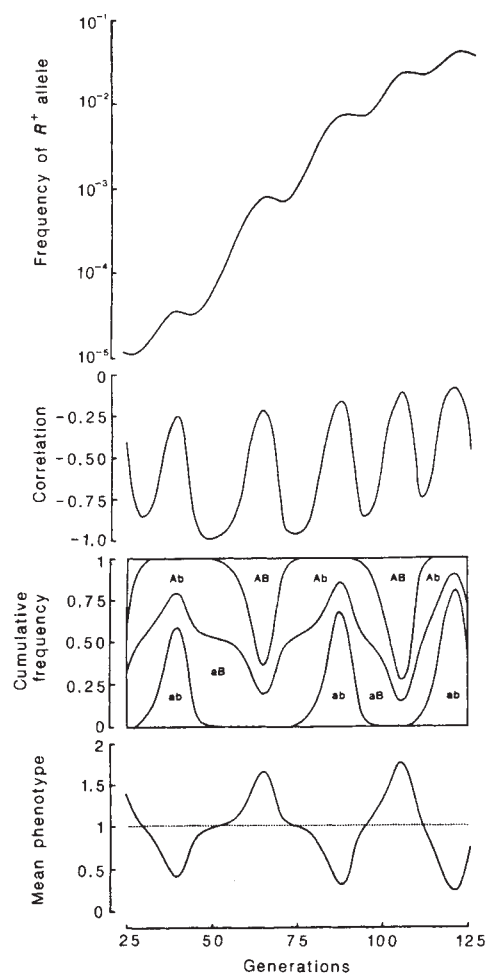


Fig. 2 Selection for recombination in a quantitative model. These diagrams illustrate changes in the 'parasite' population over 100 generations. The bottom diagram shows the oscillation of mean phenotype, with the changing genotype frequencies given above. Note that *Ab* and *aB* are both always abundant, with selection alternately favouring *ab* and *AB*. The next diagram (second from top) shows how this flux of genotypes results in an oscillation of the correlation between loci (expressed as the correlation between genotypic values, to avoid the dependence on gene frequency of the coefficient of linkage disequilibrium). Finally, the top diagram shows the correlated spread of the R^+ gene, from its initial frequency of 10^{-5} to a final frequency of >0.25 . The spread of R^+ is much more regular and more rapid than in the gene-for-gene model.

Simulation of a quantitative model is shown in Fig. 2. In this model, the performance of a given parasite infesting a given host type is determined by the absolute value of the difference between their phenotypes, with *ab* having a phenotype of 0, *aB* and *Ab* of 1, and *AB* of 2. Its performance is greatest when infesting a host with the same phenotype. The performance of a host type is likewise determined by the absolute difference between its phenotype and that of the parasite type with which it is interacting, but host susceptibility is least when this difference is greatest. These assumptions satisfy the conditions for a cyclical game¹⁰, and create a divergent oscillation of mean phenotype in both host and parasite. The R^+ allele always spreads in the parasite and is lost from the host (we emphasize that 'host' and 'parasite' refer, respectively, to the species that is fittest when most different from its antagonist, and to the species that is fittest when most similar). The response matrix of the pathogen implies that selection is generally stabilizing, so that most of the time there is an excess of *Ab* and *aB* individuals, whose phenotype is intermediate. It is therefore

often the case that the population consists mainly of a more or less equal mixture of aB and Ab , while selection is favouring a phenotype of zero (or 2) and thus ab (or AB). Selection at these times favours a change in the sign of the linkage disequilibrium, and the R^+ allele increases in frequency, essentially because the progeny of R^+ parents are closer to linkage equilibrium than the majority of the population. In the host population, selection alternately favours the two extreme genotypes ab and AB . The population comes to consist almost entirely of these two genotypes. There is no selection for a change in the sign of the linkage disequilibrium, and the frequency of the R^+ allele declines to zero. These conclusions, that R^+ decreases in the host and increases in the parasite, are insensitive to the intensity of selection.

The results of the gene-for-gene simulation were highly irregular. If only a single fitness locus is segregating, the effect of the evolutionary lag in the response of one species to another is to cause a divergent oscillation of gene frequency. When both loci are segregating simultaneously, selection at one locus interferes with selection at the other, to create irregular fluctuations in gene frequency. Both populations show rapid shifts from being almost fixed for one genotype to being almost fixed for another, with the sequence of genotypes not obeying any regular pattern. These changes are accompanied by changes in the frequency of the R^+ alleles. The initial trend is upwards, in both host and parasite populations. (The reason for this increase is that from time to time either population may include both Ab and aB chromosomes while selection is favouring ab (or AB). The R^+ gene then spreads by hitching a ride on the favoured coupling chromosome which it creates by recombination between Ab and aB .) In replicate runs with slightly varying initial conditions, the R^+ allele increased in both species from an initial frequency of 10^{-5} , but at a very variable rate. After about 2,000 generations, a value of approximately 250×10^{-5} was reached, and this was maintained in all replicates, with minor fluctuations but with no general trend, for a further 8,000 generations.

In the quantitative model, then, the high recombination allele, R^+ , increases in frequency in one species, the parasite, and decreases in the other, the 'host'. An explanation is that in the parasite there is stabilising selection for a fluctuating optimum, whereas in the host selection is disruptive. In the gene-for-gene model, R^+ alleles spread in both species, although the increase is slow and irregular. These conclusions were based on simulations in which the initial frequency of the R^+ alleles was low (10^{-5}), and in which there was no recombination in zygotes homozygous for the R^- alleles. The conclusions of the quantitative model continue to hold if the initial frequency of R^+ is 0.6 in both species, and if recombination in the R^- homozygote was 0.05 between R and A , and between A and B . However, if either of these modifications are introduced into the gene-for-gene model, the general tendency for the R^+ alleles to increase in frequency disappears.

These models confirm that negative species interactions may be central to the evolution of phenomena such as sexuality, recombination and outcrossing, which ensure that a previously successful strategy is systematically abandoned. Most of the comparative data are consistent with this theory⁸, in showing that sexual reproduction is most likely to be lost in environments with relatively few species.

Comparative evidence supports the Hamilton and Zuk hypothesis on parasites and sexual selection

Andrew F. Read

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Darwin¹ proposed that secondary sexual characters, such as the long tails and bright plumage of many birds, evolved because females use them as cues in male choice. The question of why females should prefer males with such apparently deleterious characters is currently the subject of vigorous debate²⁻⁷. Hamilton and Zuk⁸ suggest that females use secondary sexual characters to assess a male's ability to resist parasites. A prediction of this hypothesis is that male brightness should correlate positively with parasite load across species, and the only evidence advanced in support of the model is Hamilton and Zuk's⁸ finding that such a correlation exists across North American passerines. But inter-specific correlations of this sort can result from phylogenetic associations through common descent or from independent associations with some confounding variable, such as an aspect of behaviour or ecology⁹⁻¹³. Here, I use data on European passerines and an enlarged data set on North American passerines to demonstrate positive relationships between male brightness and parasite prevalence which remain when the effects of taxonomic, behavioural and ecological variables are removed.

Hamilton and Zuk⁸ have proposed that females use colour and other secondary sexual characters to assess a male's ability to resist parasites (defined here to include parasitic viruses, bacteria, protozoa and helminths): males which are more resistant to parasites are assumed to be brighter (or have longer tails) and, if resistance is heritable, a female will increase the viability of her offspring by mating with brighter males. Thus, in species which are particularly susceptible to parasite invasion (and therefore in which variation in parasite resistance is a major heritable component of variation in fitness), sexual selection should favour characters which allow females to judge a male's present and past parasite load. Using parasite prevalence as a measure of a species' susceptibility to infestation, the model therefore predicts that, across species, the extent to which such characters are expressed should correlate positively with parasite prevalence. In a preliminary analysis, Hamilton and Zuk found a positive association between brightness and the prevalence of avian haematzoa across 109 species of North American passerines. But there were two major problems with their analysis. First, the correlation could have been due to groups of related species sharing the same traits, not because of convergent evolution, but because of their shared ancestry⁹⁻¹³. Second, the positive association between brightness and parasites may exist only because of other variables which are independently associated with both parasite prevalence and brightness¹²⁻¹⁴. Here, I use large data sets on the prevalence of avian haematzoa in both North American and European passerines to test the Hamilton and Zuk model after the effects of taxonomy and other potentially confounding variables have been removed.

The colour of each sex of each passerine species during breeding was scored on a six point scale of brightness, with the brightest birds given the highest scores. I used Hamilton and Zuk's⁸ brightness scores for the North American species and, for the European birds, brightness ratings used by Baker and Parker¹⁵ in their analysis of bird coloration. Baker and Parker's rankings were determined before the sexual selection and parasite hypothesis had been proposed. They scored seven regions of a bird's body for colour and I averaged these to give a brightness ranking for each species and sex. Data on parasite infestations were taken from surveys of the avian haematzoa

Received 29 January; accepted 13 May 1987.

1. Maynard Smith, J. *J. theor. Biol.* **30**, 319-335 (1971).
2. Charlesworth, B. *Genetics* **83**, 181-195 (1976).
3. Jaenike, J. *Evol. Theory* **3**, 191-194 (1978).
4. Hamilton, W. D. *Oikos* **35**, 282-290 (1980).
5. Tooby, J. *J. theor. Biol.* **97**, 557-576 (1982).
6. May, R. M. & Anderson, R. M. *Proc. R. Soc. B* **219**, 281-313 (1983).
7. Hutson, V. & Law, R. *Proc. R. Soc. B* **213**, 345-359 (1981).
8. Bell, G. *The Masterpiece of Nature* (Croom Helm, London, 1982).
9. Flor, H. H. *Adv. Genet.* **8**, 29-54 (1956).
10. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge University Press, 1982).

Table 1 Associations between the prevalence of blood parasites and brightness across North American and European passerines

	Male brightness		Female brightness	
	r_s	P	r_s	P
North American passerines ($n = 114$)	0.256	0.003	0.175	0.031
European passerines ($n = 113$)	0.202	0.016	0.095	0.156

r_s = Spearman rank correlation coefficient, P = probability value.

of wild birds caught throughout the two zoogeographic regions, and includes blood samples taken from 23,624 individuals in North America, and from 10,463 individuals in Europe¹⁶⁻¹⁸. Infestation levels for each bird species were calculated as the proportion of sampled individuals which were infected with any blood parasites.

There is a positive and significant correlation between the prevalence of blood parasites and brightness across male and female North American and male European passerines (Table 1), as predicted by the Hamilton and Zuk hypothesis. But these associations could be the result of a spurious coincidence in one or a few speciose taxa. If the positive correlation between colour and parasite prevalence across passerines really does exist, then it should have arisen many times during the course of evolution. Thus, the Hamilton and Zuk hypothesis predicts that *within* genera brighter species should have larger parasite loads.

This prediction was tested by calculating a Spearman rank correlation coefficient (r_s) for every genus containing two or more species. Randomization tests^{11,19} show that the within genus correlations between male brightness and parasite prevalence are more positive than expected by chance alone (Table 2). The mean correlation coefficients in the two regions are

Table 2 Association between the prevalence of blood parasites and brightness within genera

	Male brightness			Female brightness		
	mean r_s	P	n	mean r_s	P	n
North American passerines	0.37	0.036	15	0.04	0.44	14
European passerines	0.28	0.088	22	0.12	0.235	23
Pooled	0.33	0.008	32	0.09	0.283	33

Five genera were common to both the North American and European samples and their correlation coefficients were averaged in the pooled analysis. Probability values were calculated from a randomization test: the sign of each generic correlation coefficient was randomly changed with a probability of 0.5, and the mean of these randomized coefficients calculated. This was repeated 10,000 times, and the proportion of times that the mean of the randomized coefficients was greater than the mean of the observed coefficients was taken as the probability value. Mean r_s = mean generic correlation coefficient, n = number of genera; other abbreviation as in Table 1.

positive, and the correlations of the American genera are significantly greater than zero. The positive effect within the European genera is not significant, but the probability value is suggestive. And when the intrageneric correlations within both geographic regions are pooled, they are significantly greater than zero ($0.01 > P > 0.001$), with 22 of 32 genera exhibiting positive rather than negative correlations. Thus the positive associations between male brightness and parasite prevalence found across species cannot be due to taxonomic artifacts alone. The mean correlation coefficients between female brightness and parasite prevalence within genera are positive, but not significant in either of the regions or in the pooled sample. The remainder of the analysis deals only with male brightness.

Many behavioural, ecological and morphological variables

Table 3 Association between the prevalence of blood parasites and male brightness among European genera when the effects of latitudinal variables are removed

Variable held constant	Mean generic r_s	P	n
Breeding range	0.32	0.023	24
Wintering range	0.33	0.016	24
Distance migrated	0.29	0.034	24

The breeding and wintering latitudes for a species were defined as the mean of the northern-most point in the Western Palearctic and the southern-most latitude in the Northern Hemisphere of the breeding and wintering ranges; distance migrated is the latitudinal distance between the mean breeding and wintering latitudes. Abbreviations as in Tables 1 and 2.

could be responsible for the correlations between parasite prevalence and brightness. The following are all the suggested^{1,15} correlates of bird coloration known to me: body size, nest type, relative effort by the male in incubation and nestling care, diet, habitat, breeding latitude, winter latitude, distance migrated, mating system, nest dispersion, activity timing and feeding gregariousness during winter. Apart from body length, data on these variables for many North American species are not readily available, so the analysis of possible confounding variables was restricted to European passerines. The definition and categorisation of variables, and the data, come largely from Baker and Parker¹⁵ and Bennett²⁰; details will be given elsewhere.

The effects of continuous variables were removed by calculating the intrageneric rank correlations between deviations from best fit lines fitted across all species by least squares regression. When the effect of body size is removed in this way (using the average distance from the tip of the bill to the tip of the tail as a measure of size), the intrageneric associations between parasite prevalence and colour remain significantly greater than zero in North America, and become significant in Europe (North America: mean generic $r_s = 0.37$, $P = 0.036$, $n = 15$; European: mean generic $r_s = 0.31$, $P = 0.036$, $n = 23$; pooled sample: mean generic $r_s = 0.33$, $P = 0.006$, $n = 32$). Similarly, in the European birds, the intrageneric associations become significantly positive when the effects of breeding latitude, wintering latitude, and latitudinal distance migrated are removed separately (Table 3), and when the effects of body size, breeding and wintering latitudes are removed simultaneously using multiple regression (results of the randomization test: mean generic $r_s = 0.28$, $P = 0.028$, $n = 24$ genera). The fact that in all cases, the intrageneric correlations between colour and parasite prevalence in European passerines become significantly positive when other factors are held constant shows that these variables slightly obscure the parasite-colour relationship.

Many categorical variables could not be important correlates of intra-generic variation in brightness because they do not vary in most or all genera. Others do vary within a large proportion of genera (from 6 to 13 out of 22), and might therefore obscure or strengthen any associations between colour and parasites. To test whether variation in any of these factors had a significant effect on the association between colour and parasite prevalence, I compared the magnitude of the parasite-colour correlations of those genera that exhibited variation in the ecological factor in question with those in which there was no variation (Table 4). There is no evidence that the positive intra-generic associations between colour and parasite prevalence are confounded or produced by any of the ecological variables examined.

The above analysis demonstrates that, as predicted by Hamilton and Zuk⁸, parasite prevalence and male brightness are positively correlated across species: across European passerines when the effects of confounding variables are removed, and across North American passerines, parasite prevalence and brightness correlate within genera, demonstrating that the

Table 4 The effect of behavioural and ecological variables on the correlations between parasite prevalence and male brightness within European genera

Ecological variables	Proportion of genera exhibiting variation	t	P
Diurnal/nocturnal	0/22	—	—
Male nestling care	0/22	—	—
Diet	4/22	—	—
Territoriality/coloniality	4/22	—	—
Hole/open/ground nesting	7/22	0.759	0.23
Winter feeding	6/22	1.01	0.16
gregariousness			
Male incubation effort	10/22	1.154	0.13
Mating system	13/19	0.444	0.33
Habitat type	12/22	0.128	0.45
Open/forested habitat	9/22	0.829	0.21

The first four variables could not have an important effect because in all or the large majority of genera they did not vary. For the remaining variables, one-tailed unpaired *t*-tests were used to compare the magnitude of the correlations between parasite prevalence and male brightness in genera exhibiting variation in a variable with those in which there is no variation. In each case there was no significant effect. Open/forested habitat is a two-category classification of habitats in which only wooded and unwooded habitats are distinguished; habitat type refers to an eight-category classification (after ref. 20). Data on the mating systems of some species within three genera were unavailable and these genera were dropped from the analysis. Abbreviations as in Tables 1 and 2.

association has arisen many times during the course of evolution. In addition, there is no evidence that any previously suggested correlate of male brightness explains the intra-generic associations. These conclusions are all the more surprising given my methodology and the nature of the data: (1) it might be expected that any associations within genera would be obscured by sampling error because of the small number of species in each genus; (2) the subjectivity of the colour data is likely to introduce variation that will obscure rather than produce associations between brightness and parasite loads (assuming that the colour ratings have been derived independently of knowledge of the parasite loads which is clearly the case for the European data); (3) the measure of a species' susceptibility to parasite infection is not particularly accurate because I have been unable to calculate the mean parasite burden for each species, and in many cases the species' prevalence measures were based on only a few sampled individuals; and (4) I have ignored other parasite groups, such as viruses and bacteria, which could be involved⁸. But despite all these limitations, the novel prediction by Hamilton and Zuk⁸ of a positive association between brightness and parasite prevalence across species is borne out by the comparative data.

I thank R. R. Baker and G. A. Parker, W. D. Hamilton and M. Zuk and P. M. Bennett for access to their unpublished data, Paul Harvey, Alan Grafen (who suggested looking within taxa), and Mark Pagel for constructive criticism throughout, and M. A. Elgar, H. C. J. Godfray, T. Guilford, W. D. Hamilton, A. E. Keymer, and P. J. Weatherhead for useful comments. I am supported by a Commonwealth Scholarship from the Association of Commonwealth Universities and the British Council.

Received 23 March; accepted 7 April 1987.

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, London, 1871).
2. Partridge, L. & Harvey, P. H. *Nature* **323**, 580–581 (1986).
3. Andersson, M. *Evolution* **40**, 804–816 (1986).
4. Lande, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3721–3725 (1981).
5. Kodric-Brown, A. & Brown, J. H. *Am. Nat.* **124**, 309–323 (1984).
6. Kirkpatrick, M. *Evolution* **36**, 1–12 (1982).
7. Maynard Smith, J. *J. theor. Biol.* **115**, 1–8 (1985).
8. Hamilton, W. D. & Zuk, M. *Science* **218**, 384–387 (1982).
9. Harvey, P. H. & Mace, G. M. in *Current Problems in Sociobiology* 343–362 (Cambridge University Press, 1982).
10. Harvey, P. H. & Clutton-Brock, T. H. *Evolution* **39**, 559–581 (1985).

11. Elgar, M. A. & Harvey, P. H. *Functional Ecology* **1**, 23–36 (1987).
12. Clutton-Brock, T. H. & Harvey, P. H. in *Behavioural Ecology. An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 7–29 (Blackwell, Oxford, 1984).
13. Ridley, M. *The Explanation of Organic Diversity* (Clarendon, Oxford, 1983).
14. Borgia, G. *Behav. Ecol. Sociobiol.* **19**, 355–358 (1986).
15. Baker, R. R. & Parker, G. A. *Phil. Trans. R. Soc. Lond. B* **287**, 63–130 (1979).
16. Peirce, M. A. *J. natur. Hist.* **15**, 419–458 (1981).
17. Griener, E. C., Bennett, G. F., White, E. M. & Coombs, R. F. *Can. J. Zool.* **53**, 1762–1787 (1975).
18. Collins, W. E., Jeffery, G. M., Skinner, J. C., Harrison, A. J. & Arnold, F. J. *Parasitol.* **52**, 671–673 (1966).
19. Sokal, R. R. & Rohlf, F. J. *Biometry* 787–795 (Freeman, New York, 1981).
20. Bennett, P. M. *thesis*, Univ. of Sussex (1986).

Long-term potentiation of synaptic transmission in the hippocampus induced by a bee venom peptide

E. Cherubini*, Y. Ben Ari*, M. Gho*, J. N. Bidard† & M. Lazdunski†

* INSERM U29, Hôpital de Port-Royal, 123 Bd de Port-Royal, 75014 Paris, France

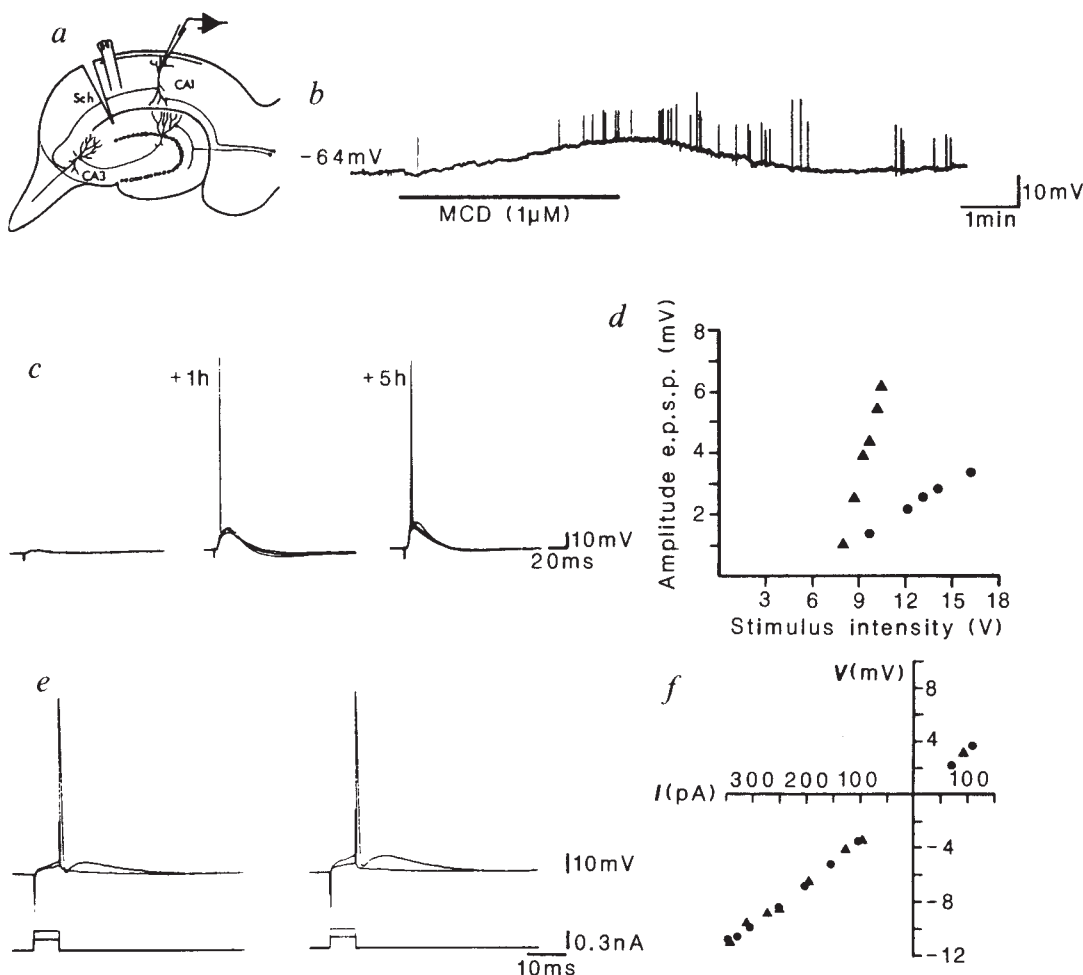
† Centre de Biochimie du CNRS, Parc Valrose, 06034 Nice Cedex, France

Several neurotoxins have been isolated from bee venom¹. One of these, the mast cell degranulating peptide (MCD), releases histamine from mast cells and on central administration produces arousal at low concentrations and convulsions at higher doses^{2,3}. These effects are mediated through specific high-affinity binding sites⁴ which are concentrated in cortical structures, notably the hippocampus⁵. This structure appears to be the source of changes in the electrocorticogram that follow injections of MCD into the cerebral ventricle, and which induce a quasi-permanent hippocampal theta rhythm in the motionless rat alternating with epileptiform spike waves⁶. We report here that brief application of MCD to the CA1 region of hippocampal slices induces long-term potentiation, that is, a long-lasting increase in the efficacy of synaptic transmission. This potentiation seems to be indistinguishable from the classical LTP produced by trains of high-frequency electrical stimulation^{5,6} and considered to be related in some way to memory. Using binding to synaptosomal membranes and radioimmunoassay techniques, we have also found an endogenous peptide equivalent of MCD in brain extracts. This raises the possibility that a MCD-like peptide may be important in long-term potentiation.

Conventional hippocampal slices were prepared from adult male wistar rats and maintained *in vitro* (details in ref. 7). Application of MCD (0.5–2 μ M) produced a membrane depolarization, often associated with spike discharge (Fig. 1b). MCD (1 μ M) produced a depolarization of 12.5 ± 1.28 mV, $n = 6$ (mean $\pm$ s.e.m.). The depolarization started 60 s after MCD application and reached a peak within a few minutes. The membrane potential returned to the control level within 3–5 min of washing with MCD-free solution. The depolarization was usually not associated with a significant change in input resistance; a slight (<10%) increase was observed in two cases. Both tetrodotoxin (TTX) (1 μ M) and cobalt (2 mM) completely blocked the depolarization produced by MCD, implying that this effect is synaptically mediated.

MCD consistently produced a long-term potentiation (LTP) of synaptic transmission ($n = 6$). This was characterized by a progressive increase of the amplitude of the excitatory postsynaptic potential (e.p.s.p.) starting 2–5 min after the wash and reaching a maximum 10–20 min later. Often the facilitated e.p.s.p. reached the spike threshold (Fig. 1c). This enhanced e.p.s.p. lasted for up to 6 h (the longest duration of stable intracellular recordings which was obtained) and even partial recovery was not observed. The enhancement produced by 1 μ M

Fig. 1 MCD enhances synaptic transmission in the hippocampus. *a*, Experimental arrangement: intracellular recordings of the pyramidal cells of CA1 were made with potassium-acetate-containing electrodes; bipolar stimulating electrodes were positioned in stratum radiatum to activate the Schaffer collaterals. The CA1 region was isolated from CA3 by a knife cut to prevent spread of bursting activity from the latter region²⁰. *b*–*e*, The effects of MCD in the same neuron. *b*, Depolarization and spike activity produced by bath application of 1 μ M MCD (bar). *c*, Superimposed digitized chart records of monosynaptic e.p.s.p. before, 1 h and 5 h after MCD. *d*, Plot of the amplitude of the e.p.s.p. against stimulation intensity before (●) and after (▲) MCD. *e*, Action potentials directly evoked by intracellular depolarizing pulses before (left) and after (right) MCD. The excitability of the neuron was not evidently changed by the peptide. Upper trace, voltage recordings; lower trace, current recordings. *f*, *V*/*I* curve constructed before (●) and 1 h after (▲) application of MCD. In this and following figures the frequency of stimulation of the Schaffer collaterals was 0.05 Hz.



Methods. MCD was purified according to Taylor *et al.*⁴, and the MCD II fraction was used throughout. The purity of the peptide was improved by an additional step of HPLC purification on C18 columns (10 $\times$ 250 mm, 7 μ m particle size, Merck). Elution was performed with an acetonitrile water gradient containing 0.5% trifluoroacetic acid and 0.85% triethylamine. MCD was dissolved in water. Because the peptide is positively charged, plastic and glass tubes were treated with Sigmacoat (Sigma).

of MCD measured in three cases in which spikes were not produced was $259 \pm 33\%$ (mean $\pm$ s.e.m.). The relationship between the amplitude of the e.p.s.p. and the stimulation intensity was linear before and after application of MCD but the slope of the responses obtained after superfusion of the toxin was considerably increased (Fig. 1*d*). The minimum dose of MCD required to produce the potentiation was 500 nM, but higher concentrations (1–2 μ M) produced larger and more rapid effects. The enhancement of the e.p.s.p. was not associated with changes in membrane resistance or cellular excitability as measured with intracellular current pulses (Fig. 1*e–f*). Therefore, the potentiation of synaptic transmission produced by MCD, like that induced by electrical stimulation⁶, does not seem to involve long-lasting changes in postsynaptic cell excitability.

The long-lasting effects of MCD are not due to the continuous presence of the toxin in the tissue because the effects of MCD on membrane potential and resistance rapidly washed out and a brief (1–3 min) concomitant exposure to TTX or cobalt (2 mM) completely prevented the long-lasting effects of MCD ($n=2$). Interestingly, once the potentiation had been induced, brief application of either agent caused only a temporary block of the long-lasting effects ($n=3$). This suggests that the potentiating effect of MCD requires synaptic activity.

To examine whether the effects of MCD were due to a fragment of the peptide, we also tested the effects of MCD pre-

treated either with chymotrypsin, which does not destroy the MCD peptide, or with trypsin, which does. The long-lasting enhancement of the e.p.s.p. still occurred with the former treatment ($n=2$) but not with the latter ($n=3$) which confirms the need for the peptide to be intact for activity.

Blockade of Cl^- -mediated GABAergic (γ -amino butyric acid) mediated inhibition does not prevent the development of LTP produced by trains of electrical stimulation⁸. We have therefore examined the effects of MCD in the presence of the GABA antagonist bicuculline. To suppress the synchronous firing which occurs in these conditions⁹, high concentrations of divalent cations were used (6 mM Mg^{2+} instead of 1.3 mM, and 4 mM Ca^{2+} instead of 2 mM). MCD (2 μ M) induced long-term potentiation of synaptic transmission but not the initial depolarization (Fig. 2; $n=2$). We conclude that GABAergic inhibition is not important in the long-lasting enhancement of the e.p.s.p. induced by MCD.

Several mechanisms could account for the enhancement of the e.p.s.p., including an increase in the number of presynaptic fibres excited by the electrical stimulation. To test this possibility we have examined the effects of MCD upon the afferent volleys and the field e.p.s.ps. The curve relating the stimulation intensity and the amplitude of the afferent volley was identical in control conditions and following MCD application (Fig. 3*a*). On the other hand, there was a significant shift in the input-output

Fig. 2 The enhancement of synaptic transmission by MCD is independent of membrane depolarization. *a-c*, Same neuron. *a*, Slice bathed in a solution containing 6 mM Mg^{2+} and 4 mM Ca^{2+} to reduce excitability and bicuculline (30 μM) to block GABAergic inhibition. MCD (2 μM) was applied for 5 min, only the end of the period of superfusion is represented by the bar. The peptide induces neither a depolarization nor a change in input resistance (as tested by electrotonic potentials induced by hyperpolarizing constant current pulses through the recording electrode). Scale bars: (upper) 10 mV; (lower) 0.25 nV vertical, 1 min horizontal. *b*, Superimposed digitized chart records of e.p.s.p. evoked before (left) and 2 h after (right) application of MCD. Scale bar: 10 mV vertical; 40 ms horizontal. *c*, Action potentials evoked by depolarizing current pulses before (left) and 2 h after (right) application of MCD. Scale bars: (upper) 10 mV; (lower) 0.25 nV vertical, 20 ms horizontal.

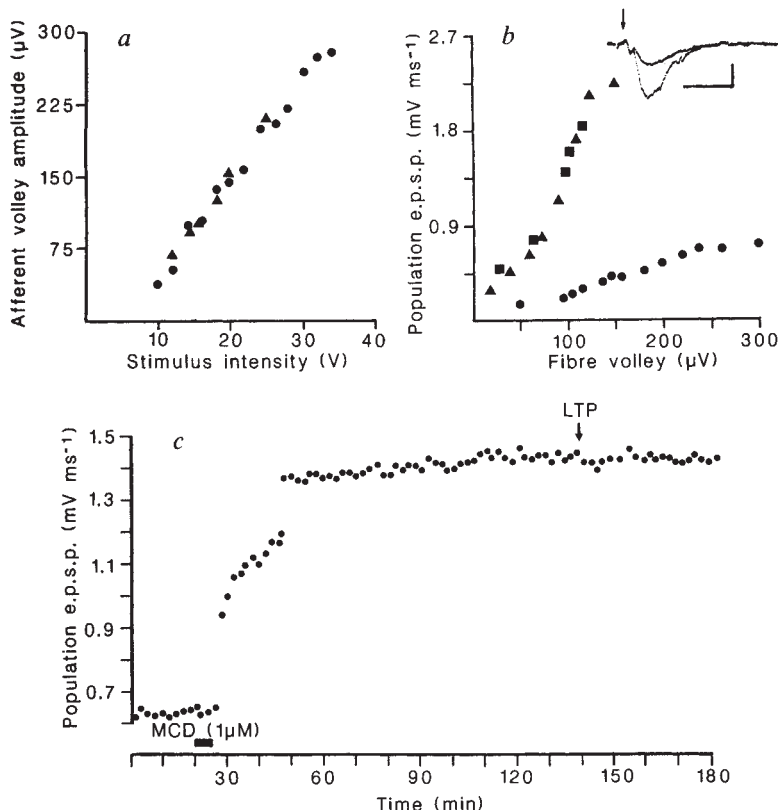
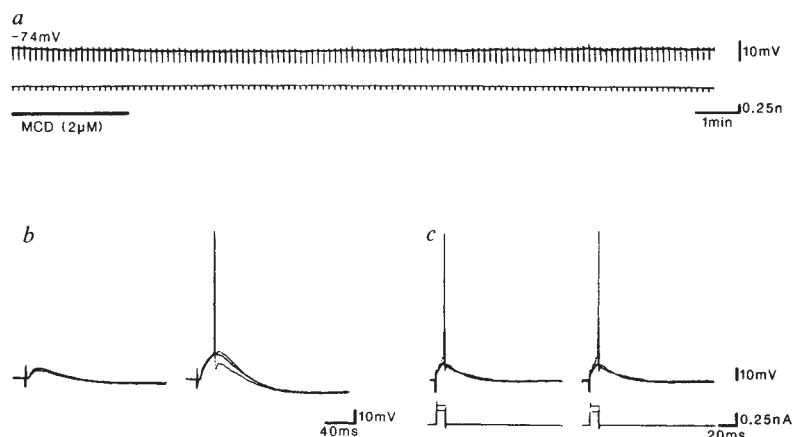


Fig. 3 *a*, The enhancement of the field e.p.s.p. produced by MCD is not associated with a change in the afferent volley. The amplitude of the afferent volley is plotted against the intensity of stimulation before (●) and 1 h after (▲) MCD (1 μM). *b*, Same slice as *a*. Input-output curve before (●) and 1 h after (▲) MCD. Electrical stimulation was repeated at this time and 30 min later (■). Each point is the average of five consecutive responses. Inset, typical digitized field e.p.s.p. and afferent volleys obtained in the control conditions and 1 h after MCD application (lower trace). Scale bar, 0.5 mV vertical, 10 ms horizontal. *c*, Time-course of the enhancement of the field e.p.s.p. produced by 1 μM MCD (bar) in another slice. Each point is the average of five consecutive responses. Electrical tetanization was induced 2 h after MCD (arrow). This did not produce a further potentiation of the synaptic response. In three other experiments LTP was first induced by trains of electrical stimulation and MCD applied 1 h later. The electrically potentiated response was further enhanced by the peptide ($175 \pm 10\%$, mean $\pm$ s.e.m.). **Methods.** The initial slope of the field e.p.s.p. was measured as an indication of the amplitude of synaptic currents⁶. LTP was induced by two trains of 100-Hz stimulation at 30-s interval. Two experimental approaches have been used to study the interactions between MCD-induced synaptic potentiation and electrical LTP. In both cases the tetanus was induced once the MCD effects were maximal (usually 1 h after application). In the first case (see *b*) the strength of stimulation used to elicit the LTP was reduced by 50% to obtain a field e.p.s.p. similar to the control (pre-drug) conditions. In the second approach (in *c*) the strength of stimulation was kept constant throughout the experiment.

curve obtained by plotting the amplitude of the afferent volley against the initial slope of the field e.p.s.p. (Fig. 3*b*). Clearly a given afferent volley produces a larger postsynaptic response after MCD, suggesting that the potentiation of synaptic transmission is not due to an enhancement of presynaptic axon excitability. A similar observation has been reported with electrically-induced LTP⁶.

Therefore the potentiation produced by MCD is strikingly reminiscent of the LTP induced by trains of high-frequency electrical stimulation^{5,6}. To examine this similarity further, trains of electrical stimulation which produce LTP in control slices were applied in MCD-pretreated slices. The trains were found to produce a frequency potentiation lasting 10–15 min (in 4 out of 6 tests) but did not produce a long-lasting enhancement of the e.p.s.p. (Fig. 3*c*). The fact that the chemically potentiated synapse cannot be further enhanced by electrical stimuli which otherwise produce LTP suggests that the enhancement of synaptic transmission by both procedures may have a common mechanism. To examine further the similarity between MCD-induced long-lasting enhancement of the synaptic transmission and LTP, we tested the effects of D(-)-2-amino-5-phos-

phonovaleric acid (APV), known to prevent LTP¹⁰. Application of APV (10–30 μM) before MCD suppressed the enhancement of the e.p.s.p. induced by MCD in three out of four cells tested.

The precise mechanism underlying the action of MCD in the hippocampus is unknown. One possibility is that, like phorbol esters, MCD activates protein kinase C (PKC). Following activation of this sort, K^+ channels could be phosphorylated and partly or completely blocked^{11–13} and/or Ca^{2+} channels activated¹⁴. Recent studies suggest that phosphorylation of proteins by PKC is important in LTP¹⁵ and that phorbol esters produce long-term potentiation in the hippocampus¹⁶. Another possibility is that MCD directly blocks or activates ionic channels involved in neurotransmitter release in the hippocampus. Another bee venom toxin, apamin, is a selective blocker of one type of neuronal Ca^{2+} -activated K^+ channel^{17,18}. Similarly to apamin, for which there is an endogenous equivalent in the central nervous system¹⁹, we have demonstrated the presence of an endogenous MCD-like activity in mammalian brain (Fig. 4). The endogenous peptide prevents ¹²⁵I-labelled MCD binding to its synaptosomal receptor and cross-reacts in a radio-immunoassay against MCD. The substance is a peptide that is

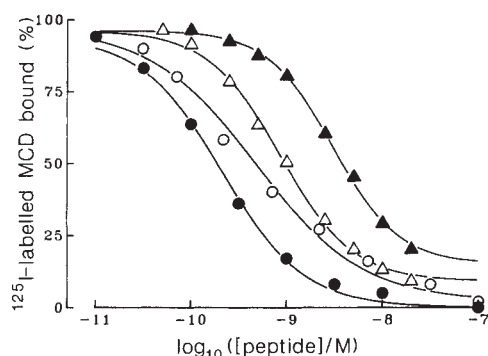


Fig. 4 Evidence for the existence of an endogenous brain substance with MCD-like activity. Effect of MCD (circles) and the MCD-like peptide (triangles) on binding of ^{125}I -MCD (4 pM) to synaptosomal membranes (0.2 mg protein ml^{-1}) in a radio-receptor assay (RRA) (filled symbols) and of the same proteins on ^{125}I -labelled MCD (8 pM) precipitation by rabbit anti-MCD immunoglobulin (final dilution 1:200,000), in a radio-immunoassay (RIA) (open symbols).

Methods. RRA was carried out as previously described⁴. Anti-MCD immunoglobulin was prepared from the serum of a rabbit immunized with MCD coupled to BSA and RIA for MCD was performed as previously described for the RIA for apamin²¹. Protamine (100 $\mu\text{g ml}^{-1}$) was added to the RIA incubation buffer to avoid nonspecific precipitation of ^{125}I -labelled MCD. The MCD-like peptide was extracted from 250 pig brains. The first steps of purification were carried out as previously described for the apamin-like substance¹⁹. Three HPLC steps were then used: one step on C_{18} -reverse phase eluted by an acetonitrile/water gradient containing 0.5% trifluoroacetic acid and 0.85% triethylamine, then two steps on TSK-IX SP-5PW cationic exchanger eluted by a 500 mM to 1,500 mM ammonium acetate gradient and finally one step on TSK G-2000 SW gel filtration eluted by 300 mM ammonium acetate pH 4.5 to yield 60 pmol of the pure MCD-like peptide (relative molecular mass = 2,900).

destroyed by trypsin, like MCD itself. The endogenous substance may induce long-term potentiation in the hippocampus, a possibility that will be tested when larger amounts of the substance become available.

We thank Drs K. Krnjevic, G. Romey and C. Van Renterghem for comments on the manuscript, Dr H. Schweitz for the anti-MCD immunoglobulin and C. Roulinat-Bettelheim for technical assistance. This work was supported by the Institut National de la Santé et de la Recherche Médicale, the CNRS and the French Ministère de la Défense. Michel Gho is a fellow of the Fondation FYSSEN.

Received 20 February; accepted 23 April 1987.

- Habermann, E. *Science* **177**, 314-322 (1972).
- Habermann, E. *Naunyn-Schmiedeberg's Archs Pharmacol.* **300**, 189-191 (1977).
- Bidard, J. N., Gondolfo, G., Mourre, C., Gottesman, C. & Lazdunski, M. *Brain Res.* (in the press).
- Taylor, J. W., Bidard, J. N. & Lazdunski, M. *J. biol. Chem.* **259**, 13957-13967 (1984).
- Bliss, T. V. P. & Lomo, T. *J. Physiol., Lond.* **232**, 331-356 (1973).
- Andersen, P., Sundberg, S. H., Sveen, O., Swann, J. W. & Wigstrom, H. *J. Physiol., Lond.* **302**, 463-482 (1980).
- Gho, M., King, A. E., Ben Ari, Y. & Cherubini, E. *Brain Res.* **385**, 411-414 (1986).
- Wigström, H. & Gustafsson, B. *Nature* **301**, 603-604 (1983).
- Miles, R. & Wong, R. K. S. *J. Physiol., Lond.* **373**, 397-418 (1986).
- Wigström, H., Gustafsson, B. & Huang, Y. Y. *Neuroscience* **17**, 1105-1115 (1986).
- Baraban, J. M., Snyder, S. H. & Alger, B. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2538-2542 (1985).
- Malenka, R. C., Madison, D. V., Andrade, R. & Nicoll, R. A. *J. Neurosci.* **6**, 475-480 (1986).
- Kandel, E. R. & Schwartz, J. H. *Science* **218**, 433-443 (1982).
- De Riemer, S. A., Strong, J. A., Albert, K. A., Greengard, P. & Kaczmarek, L. K. *Nature* **313**, 313-316 (1985).
- Nelson, R. B. & Routtenberg, A. *Expl Neurol.* **89**, 213-224 (1985).
- Malenka, R. C., Madison, D. V. & Nicoll, R. A. *Nature* **321**, 175-177 (1986).
- Hugues, M., Romey, G., Duval, D., Vincent, J. P. & Lazdunski, M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1308-1312 (1982).
- Lazdunski, M. *Cell Calcium* **4**, 421-428 (1984).
- Fosset, M., Schmid-Antomarchi, A., Hugues, M., Romey, G. & Lazdunski, M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7228-7232 (1984).
- Schwartzkroin, P. A. & Prince, D. A. *Brain Res.* **147**, 117-130 (1978).
- Schweitz, H. & Lazdunski, M. *Toxicol.* **22**, 985-988 (1984).

Differential modulation of three separate K-conductances in hippocampal CA1 neurons by serotonin

A. Colino* & J. V. Halliwell†

MRC Neuropharmacology Research Group,
Department of Pharmacology, University of London,
29/39 Brunswick Square, London WC1N 1AX, UK

The hippocampus receives a dense serotonin-containing innervation from the divisions of the raphe nucleus^{1,2}. Serotonin applied to hippocampal neurons to mimic the action of endogenous transmitter often produces complex and variable responses (see for example ref. 3). Using voltage-clamp methods and new ligands that are selective for subtypes of serotonin receptors^{4,5}, we have been able to clarify the mechanism of serotonin action on CA1 cells in rat hippocampal slices. We describe three distinct actions of serotonin (or 5-HT) on identified K-conductances in these cells. First, it activates a Ca-independent K-current which is responsible for neuronal hyperpolarization and is inhibitory. Second, it simultaneously suppresses the slow Ca-dependent K-conductance that is largely responsible for the accommodation of cell firing in CA1 neurons⁶⁻⁸: this produces a paradoxical increase in neuronal discharge in response to a depolarizing input. Third, serotonin produces a more slowly developing and long-lasting suppression of an intrinsic voltage-dependent K-conductance, I_m (ref. 9), leading to neuronal depolarization and excitation. The hyperpolarizing response is mediated by class 1A serotonin receptors, whereas the other responses are not. Modulation of these different conductances by endogenously released serotonin could therefore change the probability or the duration (or both) of neuronal firing in the mammalian brain in different ways to give inhibitory, excitatory or mixed effects.

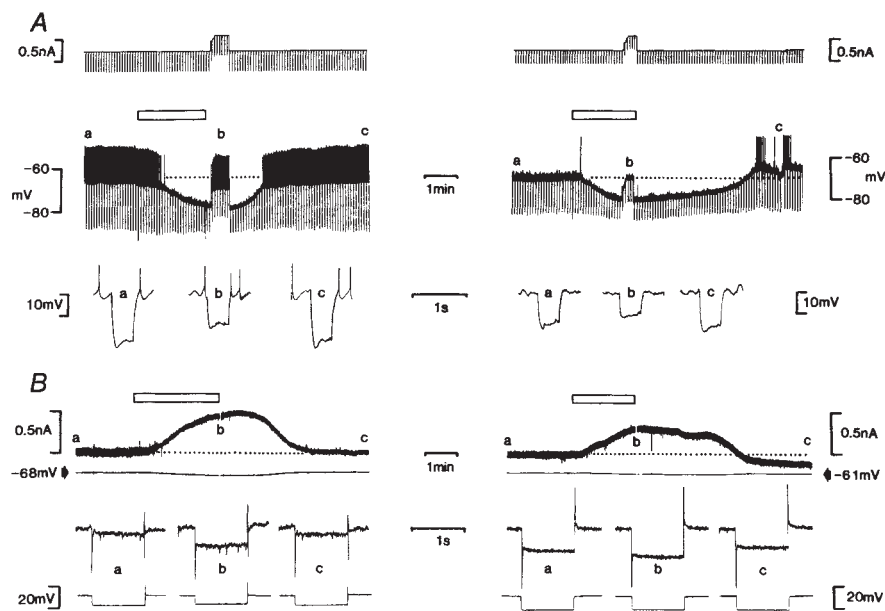
Slices were prepared from either the septal or temporal poles of the rat hippocampus, incubated and investigated electrophysiologically as previously described⁹. CA1 neurons were impaled with micropipettes containing 3 M KCl. When serotonin (10–30 μM) was bath-applied in the superfusion medium, most CA1 neurons (92%) initially hyperpolarized and showed a conductance increase that was still present when the potential shift was offset by passing positive current through the recording electrode. These effects started about 90 s after the switch to serotonin-containing medium, reached a peak after about 2–3 min in serotonin and declined 2–3 min following washout (Fig. 1A). Although no difference was noted in the passive membrane properties of neurons in septal hippocampus compared with those in cells from temporal regions, a difference between the two locations in the responses to applied serotonin was detected¹⁰. Most septal CA1 cells (70%, $n = 23$) showed a monophasic voltage change whereas the remainder and 65% of temporal cells ($n = 40$), after the hyperpolarizing phase, showed a rebound depolarization which was either slow to dissipate or failed to recover (Fig. 1A, light panel). Overall 10% of temporal neurons were slowly depolarized by serotonin whereas 22% showed only hyperpolarization. The depolarizations to serotonin were accompanied by a neuronal conductance decrease. Both phases persisted in medium containing 0.5 μM TTX (tetrodotoxin) and in medium containing Cd (100–300 μM) or low Ca plus Mn (2 mM) to block Ca channels. The hyperpolarizing phase was inhibitory and the depolarizing phase excitatory, judged in terms of spontaneous neuronal discharge (Fig. 1A).

When voltage-clamped in TTX-containing medium at near

* Present address: Departamento de Fisiología, Facultad de Medicina, Fouseca 2, 37007 Salamanca, Spain.

† To whom correspondence should be addressed: MRC Neuropharmacology Research Group, Department of Pharmacology, University College London, Gower Street, London WC1E 6BT, UK.

Fig. 1 Hippocampal neuronal responses to serotonin under current and voltage-clamp conditions. **A**, Chart records of membrane potential (middle trace) of rat hippocampal CA1 neurons contained in transverse slices taken from close to the septal (left-hand traces) or temporal (right-hand traces) poles of the hippocampus, around the time of superfusion with 30 μ M serotonin (open bar). The negative voltage excursions are in response to 375 ms hyperpolarizing constant current injections presented every 5 s (upper trace of each part); the positive deflections represent truncated action potentials. The lower traces of each part are voltage responses at a faster recorder speed, taken before (a), during (b) and following washout (c) of serotonin at the times (a–c) indicated on the continuous voltage record. While recording the fast sweeps, the membrane potential was returned, as necessary, to the control level by the passage of steady positive or negative direct current. The septal neuron was spontaneously active and was inhibited by serotonin, whereas the temporal neuron was quiescent, but after the hyperpolarizing phase, showed rebound excitation. **B**, Voltage-clamp records from the same neurons as in **A**. In each panel there are two pairs of traces: the upper trace of each pair is a record of current, the lower trace of voltage.



The upper pair of records depict the time-course of the currents induced by application of 30 μ M serotonin (open bar) in the septal (left-hand panel) and the temporal (right-hand panel) neurons when clamped at the indicated membrane potentials. Before (a), during (b) and after (c) exposure to serotonin, the cells were subject to 1-s hyperpolarizing voltage-clamp steps. The septal neuron was studied under the same incubation conditions as in **A**; the temporal neuron was bathed in medium that also contained 0.5 μ M TTX.

Methods. Hippocampal slices (500- μ m thick) were prepared as previously described⁹ by chopping freshly-dissected rat hippocampi transversely to the septo-temporal axis. Chopping started at either the septal or temporal pole and slices from the initial one-third of the structure were retained and designated 'septal' or 'temporal' hippocampus. These slices were incubated at room temperature in oxygenated, modified Krebs' medium until required for recording (>1 h after cutting) whereupon they were transferred singly to a recording chamber and superfused (5 ml min⁻¹) with oxygenated medium at 28–30 °C. CA1 neurons were impaled with glass (1.0 or 1.2 mm shank diameter) fibre-filled microelectrodes containing 3 M KCl (tip resistance 40–80 M Ω), and recordings made either with an amplifier possessing the facility for current injection with the elimination of voltage responses due to electrode resistance ('bridge-balance') or with a home-built switching (3 kHz) single-electrode voltage-clamp amplifier⁹. Drugs were applied in the superfusion medium which was composed of (mM): Na, 146; K, 3; Ca, 2.5; Mg, 1.2; H₂PO₄, 1.2; Cl, 130; HCO₃, 25; glucose, 11; pH 7.4 when equilibrated with 95% O₂/5% CO₂. A 'Ca-free' medium was sometimes employed, which was the same, but with no H₂PO₄ and additional (4 mM) MgCl₂ and CdCl₂ (100–300 μ M) or 2 mM MnCl₂ to antagonize residual Ca.

their resting potentials, septal and temporal CA1 neurons displayed shifts in the clamping current during exposure to, and washout of, serotonin that reflected voltage recording exactly. There was an initial outward current with increased membrane conductance (Fig. 1B), and after washing, a subsequent inward current (Fig. 1B, right panel) which was more pronounced at positive holding potentials and in the temporal neurons; a reduced membrane conductance accompanied this inward current. Both outward and inward components of the serotonin response persisted in Ca-free medium. The outward current reversed between -95 and -85 mV and displayed inward rectification, being much smaller at potentials positive to -60 mV (Fig. 2a, upper graph). The serotonin-induced outward current had a more positive reversal potential in raised extracellular K (Fig. 2a upper graph) and was blocked by 1–2 mM Ba. Collectively these data suggest that serotonin activates a K-conductance that is responsible for the observed hyperpolarization. This seems to have similar properties to the conductances gated by the GABA_B agonist baclofen^{11,12} and by adenosine¹³. The inward current did not reverse clearly and comprised a relatively voltage-insensitive component of 50–100 pA and a very voltage-sensitive inward current that was strongly developed at potentials positive to -55 mV (Fig. 2a, lower graph). The conductance decrease associated with serotonin-induced inward current suggests that the suppression of some outwardly-directed conductance is the generating mechanism for this phase of serotonin action. Because the cells were impaled with KCl-containing electrodes and the chloride reversal potential was well positive to rest (all spontaneous synaptic potentials being positive-going: see ref. 14), the reduction of K-conductance is implicated. Sup-

pression of the slow Ca-activated K-conductance^{7,8} (and see below) could contribute some inward current at more positive potentials, but is unlikely to account for all the voltage-sensitive serotonin-induced inward current. Because the inward current was still present when Ca-influx was prevented, it appears that serotonin inhibits a voltage-dependent K-conductance which is activated just positive to -60 mV. The M-current, (I_m , ref. 9) is such a conductance. Consistent with an effect on I_m was the reduction of the characteristic time and voltage-dependent M-current relaxations⁹ at a time when the serotonin-induced inward current was well-developed (Fig. 2b). Furthermore, the inward current was not present in medium containing 2 mM Ba which blocks I_m (ref. 9). However, a small (and variable) amount of inward current was present well negative to rest potential and out of the M-current activation range (Fig. 2a, lower graph). It is probable, therefore, that a decrease in non-specific 'leak' conductance also results from serotonin action.

A third, entirely independent action of serotonin on hippocampal neurons was observed. Serotonin reduced the long-lasting after hyperpolarizations (AHPs) that follow spike firing triggered by depolarizing current stimuli (Fig. 3a). These AHPs are caused by Ca entry^{6,15} during action potentials and prevent prolonged neuronal discharge, thereby producing accommodation of spike firing⁶. The action of serotonin was not secondary to its inhibitory action on initial spike generation, because of the increased K-conductance. On the contrary, the accommodation of cell firing was reduced by serotonin (Fig. 3a), so that more spikes were evoked by current injection in its presence. Paradoxically, prolonged firing was observed when the agonist-induced K-conductance was maximally activated (Fig. 3a, centre

Fig. 2 *a*, Voltage dependence of serotonin-induced currents. Upper graph shows voltage and ionic-concentration dependence of serotonin-induced outward current. A septal cell was stepped under voltage-clamp from -63 mV to different potentials for 1 s before and during a 5-min application of $10 \mu\text{M}$ serotonin in a medium containing 3 mM or 9 mM K. The currents produced by such voltage commands in control medium were subtracted from their counterparts in the presence of serotonin using a fixed baseline to take account of the serotonin-induced shift in holding current. The currents thus determined (I_{5-HT}) are shown plotted against voltage achieved during the clamp step in 3 mM K ($\bullet$) and 9 mM K (Δ). Note the inward rectification of I_{5-HT} and the shift of its reversal potential (from -94 to -68 mV) in raised external K. Lower graph shows voltage dependence of serotonin-induced inward current. Current-voltage (I - V) curves for I_{5-HT} are shown, determined as above, but in a temporal neuron (holding potential, -62 mV), at the peak of the outward current induced by $30 \mu\text{M}$ serotonin ($\bullet$) and 5 min after washing out serotonin, when the subsequent inward current was maximally developed ($\square$). *b*, The serotonin-induced inward current is associated with M-current reduction. To reveal an uncontaminated inward serotonin current, a temporal hippocampal slice was preincubated in medium containing $0.5 \mu\text{M}$ 8-OH-DPAT to block the outward current. The CA1 neuron was held at -46 mV and stepped to -56 mV every 5 s. Traces show current (upper) and voltage (lower) and the moment at which serotonin ($30 \mu\text{M}$) superfusion of the slice commenced (arrow). TTX ($0.5 \mu\text{M}$) was present during all the experiments depicted.

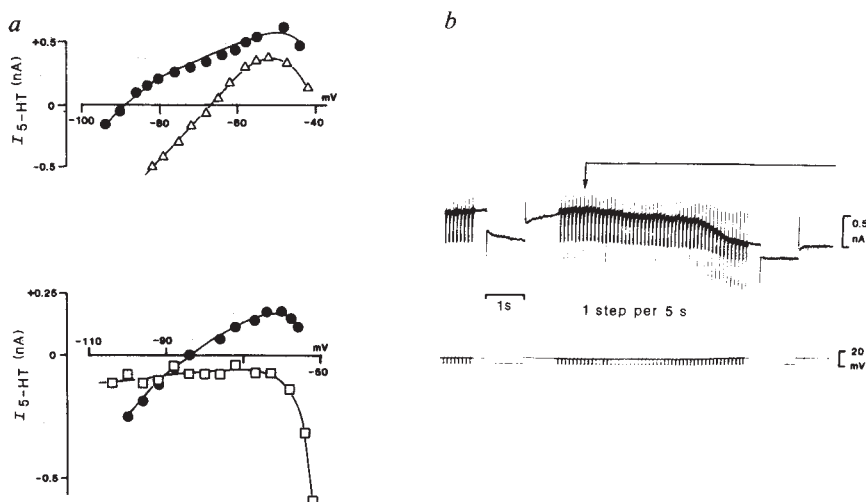
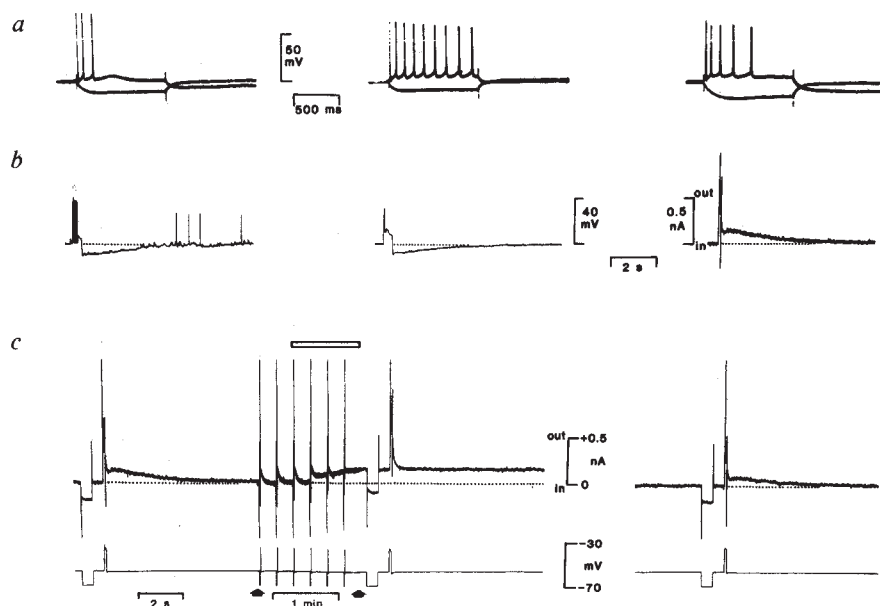


Fig. 3 Serotonin suppresses I_{AHP} and the accommodation of cell firing in the hippocampus. *a*, Oscilloscope records of superimposed voltage responses elicited in a temporal CA1 neuron when ± 0.5 nA, 375 ms current pulses were injected before (left), during (centre) and 8 min after (right) exposing the slice to $30 \mu\text{M}$ serotonin. *b*, Slower records of voltage responses to injection of 375 ms, 0.5 nA depolarizing current are shown for another temporal cell in normal medium (left) and in $0.5 \mu\text{M}$ TTX-containing medium (centre), whereas in the right-hand trace, the current elicited in the same TTX-treated neuron is depicted when the cell was briefly depolarized by 22 mV and then voltage-clamped at its resting potential (-56 mV). *c*, Left, currents (upper trace), driven by the indicated voltage steps (lower trace) in the same cell as in *b*; $30 \mu\text{M}$ serotonin was superfused as indicated (open bar). The recording speed was slowed down for the period between the arrows. Note the simultaneous induced outward current and the suppression of I_{AHP} . Right, response recorded after 15 min washout of serotonin. TTX ($0.5 \mu\text{M}$) was present in the superfusate.



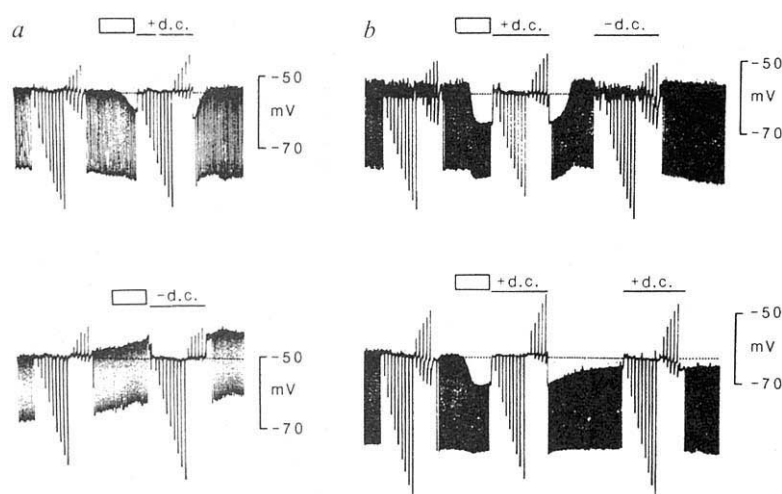
trace) (and see ref. 8) and in the case of temporal CA1 neurons, the accommodation recovered when resting-cell conductance was clearly reduced (Fig. 3*a*, right-hand trace). The actions of serotonin on accommodation were similar to the effects of other neurotransmitters, such as noradrenaline⁶, histamine¹⁶, and acetylcholine^{17,18}, as well as with activators or mimics of second-messenger systems, such as forskolin⁶, dibutyl cyclic AMP⁶ and phorbol esters^{19,20}. The AHP could be elicited in TTX-containing medium, under which conditions it was evoked by a Ca-spike (Fig. 3*b*, left and centre panels). Voltage-clamping CA1 cells at resting potential, after a brief depolarizing pulse that matched the voltage trajectory of a Ca-spike, revealed an outward current containing a long-duration component that followed the same time course as the AHP (Fig. 3*b*, right panel). This current had identical properties to the current recently described⁷ as underlying hippocampal AHPs and will be referred to as I_{AHP} . Under the more controlled conditions of the voltage-clamp, serotonin suppressed I_{AHP} (Fig. 3*c*) at a time when the

agonist-induced K-conductance was maximal. It could be argued that the loss of I_{AHP} was only apparent and that the serotonin-induced outward current reflected a persistent activation of the same Ca-dependent K-conductance with consequent occlusion of the observable response. This seems unlikely because the activation of I_{AHP} is associated with increased, not decreased, accommodation. The suppression of I_{AHP} was not mediated by an action directly upon Ca influx because serotonin did not decrease the inward current transient during depolarizing steps from rest nor did it materially depress currents through Ca channels in Ba-treated neurons (data not shown). Nor was the serotonin action mediated by enhanced spontaneous release of neurotransmitters already known to block I_{AHP} ^{6,16-18}, because serotonin still suppressed the AHP or I_{AHP} in the presence of $10 \mu\text{M}$ propranolol, $1 \mu\text{M}$ atropine or $2 \mu\text{M}$ cimetidine, which antagonized the actions of noradrenaline, acetylcholine and histamine respectively^{6,16-18}.

We attempted to identify the sites at which serotonin was

Fig. 4 Pharmacology of serotonin responses in hippocampal neurons. *a*, Voltage responses during the application of 10 μ M serotonin (box) to a septal neuron in the absence (top) and in the presence (below) of 100 nM 8-OH-DPAT, preincubated for 0.5 h. *b*, Voltage responses recorded during the application of 10 μ M serotonin (box) to a septal neuron in the absence (top) and in the presence (below) of 200 nM RU 24969 (1 h preexposure).

Methods. The voltage excursions in both *a* and *b* are in response to 375 ms, 0.5 nA hyperpolarizing current injections or 275 ms current injections of either polarity, increasing in amplitude by 0.1 nA steps, presented every 5 s. Note that, for the period when current pulses of varying amplitude were delivered, the chart speed was increased. Following voltage changes induced by serotonin, the membrane potential was returned to control level by the passage of positive or negative direct current as indicated by the lines. Action potentials were effectively filtered from the recording by employing a recorder with a slow-frequency response.



acting to produce these effects, by using compounds which were selective ligands for central serotonin (5-HT) receptors. We found no effects of the 5-HT₂ antagonist ketanserin (1 μ M, $n = 3$) on any of the responses evoked by serotonin, in agreement with the observed paucity of 5-HT₂ receptors in the hippocampus²¹. The 5-HT_{1A} ligand 8-hydroxy-2-(di-*n*-propylamino) tetralin (8-OH-DPAT) (ref. 22) did not mimic the effects of serotonin on CA1 neurons in doses of up to 10 μ M. However, 100–200 nM 8-OH-DPAT in combination with serotonin completely antagonized the hyperpolarizing action of the latter, revealing a depolarizing action of serotonin even in septal CA1 neurons (Fig. 4*a*). A concomitant block of the serotonin-induced outward current with 8-OH-DPAT (Fig. 2*b*), allowed the study of a clearer induced inward current. Another 5-HT_{1A} ligand TVX-Q 7821 (Ipsapirone)²³ (200 nM, $n = 3$) blocked the hyperpolarizing response and ($\pm$)propranolol (10 μ M, $n = 2$), which also displaces ligand binding at 5-HT₁ sites²⁴, reduced serotonin-induced hyperpolarizations by >50%. The antagonist cyproheptadine (30 μ M, $n = 7$) was also similar in action to 8-OH-DPAT. The 5-HT_{1B} ligand RU 24969 (ref. 25) (200–400 nM) and ICS 205-930, a 5-HT₃ ligand²⁶ (10 μ M, $n = 4$) both prevented the depolarizing action of serotonin in temporal CA1 neurons. In septal neurons, where few obvious serotonin-induced depolarizations occur, these two compounds significantly prolonged the hyperpolarizing action of serotonin (Fig. 4*b*). For example, in the presence of RU 24969, the half-time for the decay of the hyperpolarization elicited by 10 μ M serotonin was prolonged by $44 \pm 9\%$ ($n = 5$, $P < 0.02$; paired *t*-test). None of the compounds investigated antagonized the depressant action of serotonin on I_{AHP} .

These results show three independent actions of serotonin on CA1 hippocampal neurons. First, serotonin can activate a K-conductance, as previously suggested for hippocampal and raphe neurons^{27,28}. Second, serotonin can suppress intrinsic membrane conductance: a small component of leak (possibly non-specific) plus a K-conductance identified as I_m . The long time-course of this latter action is suggestive of an indirect effect of serotonin on membrane conductance. Third, serotonin suppresses the slow Ca-activated K-conductance I_{AHP} independently of an action on Ca-influx and in manner similar to those agents whose effects culminate in membrane protein phosphorylation^{6,16,19,20}. Serotonin effects appear to be mediated through different receptor subtypes as judged by the actions of selective ligands. The agonist-stimulated K-conductance is due to 5-HT_{1A} receptor activation (like the raphe 'autoreceptor'²⁸), because the hyperpolarizing response is antagonized by compounds which selectively displace serotonin binding at the 5-HT_{1A} site. This accords with recent observations that the inhibition of hippocampal cell population discharge produced by serotonin is antagonized by concentrations of spiperone that occupy 5HT_{1A}

sites²⁹ and that this drug also antagonizes serotonin-induced hyperpolarizations in hippocampal pyramidal neurons¹¹. Apparently, I_{AHP} is suppressed by lower concentrations of serotonin than those required to activate the agonist-induced K-conductance (J.V.H., unpublished observations) and attempts to challenge the inhibition of I_{AHP} with antagonists of other neurotransmitters known to reduce it were unsuccessful, so that it is reasonable to assume that a serotonin receptor subtype mediates this effect. Whereas ligands for the 5-HT_{1A} site blocked or reduced the agonist-activated conductance, they were ineffective against the suppression of I_{AHP} by serotonin. This provides independent pharmacological evidence that serotonin has a simultaneous and exactly opposite action on these two identified K-conductances: similar actions of noradrenaline may also occur⁸. Furthermore, 5HT_{1A} ligands often revealed a subsequent slow, depolarizing action of serotonin: therefore the reduction of intrinsic membrane conductance shown here to underlie this depolarization is also mediated by receptors other than the 5-HT_{1A} subtype. In accord with these excitatory effects of serotonin action (reduced I_m and I_{AHP}), blockade of 5-HT_{1A} sites has been shown to uncover a potentiating effect of serotonin in the hippocampus²⁹. A compound with some selectivity for the 5-HT_{1B} site, RU 24969, had, interestingly, an opposite action to that of the 5-HT_{1A}-selective ligands, namely that of prolonging serotonin-induced hyperpolarization, possibly by reducing the slowly-developing depolarization. However, until compounds with a greater selectivity for the 5HT_{1B} and other sites become available, we cannot definitively identify the receptor(s) mediating the excitatory effects of serotonin.

The three distinct actions of serotonin on hippocampal ionic conductance mechanisms described here can account for the often variable responses of mammalian central neurons to applied serotonin and clarify the physiological function of this neurotransmitter in the brain. The importance of these mechanisms in controlling cell excitability lies in the temporal overlap of their time-course and in the relative densities of the receptor subtypes that mediate their effects. The physiological consequences of serotonin, neuronally-released onto hippocampal neurons could, depending on the types of receptors activated, be inhibitory or excitatory. Activation of the K-conductance raises the firing threshold and reduces the probability of cell discharge. Suppression of I_{AHP} prolongs firing initiated by depolarizing stimuli. Suppression of I_m and 'leak' conductance is, by itself, excitatory and would furthermore potentiate the effects of additional depolarizing inputs. Possible independent control of these three mechanisms could therefore influence not only firing probability but also the tonic/phasic nature of neuronal discharge.

This work was supported by the MRC. We thank Drs David Bowen and Alan Palmer for suggesting the experiments with 8-OH-DPAT, and Dr Stephen Mitchell for help and advice.

Received 3 March; accepted 22 April 1987.

1. Moore, R. Y. & Halaris, A. E. *J. comp. Neurol.* **164**, 171-184 (1975).
2. Segal, M. & Landis, S. *Brain Res.* **78**, 1-15 (1974).
3. Jahnsen, H. *Brain Res.* **197**, 83-94 (1980).
4. Middlemiss, D. N. *Trends pharmac. Sci.* **7**, 52-53 (1986).
5. Dourish, C. T., Hutson, P. H. & Curzon, G. *Trends pharmac. Sci.* **7**, 212-214 (1986).
6. Madison, D. V. & Nicoll, R. A. *Nature* **299**, 636-638 (1982).
7. Lancaster, B. & Adams, P. R. *J. Neurophysiol.* **55**, 1268-1282 (1986).
8. Madison, D. V. & Nicoll, R. A. *J. Physiol., Lond.* **372**, 221-244 (1986).
9. Halliwell, J. V. & Adams, P. R. *Brain Res.* **250**, 71-92 (1982).
10. Colino, A. & Halliwell, J. V. *J. Physiol., Lond.* **381**, 112P (1986).
11. Andrade, R., Malenka, R. C. & Nicoll, R. A. *Science* **234**, 1261-1265 (1986).
12. Brown, D. A. & Gähwiler, B. H. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1558-1562 (1985).
13. Trussell, L. O. & Jackson, M. B. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4857-4861 (1985).
14. Alger, B. E. & Nicoll, R. A. *Brain Res.* **200**, 195-200 (1980).
15. Hotson, J. R. & Prince, D. A. *J. Neurophysiol.* **43**, 409-419 (1980).
16. Haas, H. L. & Konnerth, A. *Nature* **302**, 432-434 (1983).
17. Cole, A. E. & Nicoll, R. A. *Science* **221**, 1299-1301 (1983).
18. Benardo, L. S. & Prince, D. A. *Brain Res.* **249**, 333-344 (1982).
19. Baraban, J. M., Snyder, S. H. & Alger, B. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2538-2542 (1985).
20. Malenka, R. C., Madison, D. V., Andrade, R. & Nicoll, R. A. *J. Neurosci.* **6**, 475-480 (1986).
21. Köhler, C. *Neuroscience* **13**, 667-680 (1984).
22. Gozlan, H., El Mestikawy, S., Pichet, L., Glowinski, J. & Hamon, M. *Nature* **305**, 140-142 (1984).
23. Dompert, W. U., Glazer, T. & Traber, J. *Naunyn-Schmiedeberg's Archs Pharmac.* **328**, 467-470 (1985).
24. Middlemiss, D. N., *Eur. J. Pharmac.* **101**, 289-293 (1984).
25. Euvrard, C. & Boissier, J. R. *Eur. J. Pharmac.* **63**, 65-72 (1980).
26. Richardson, B. P., Engel, G., Donatsch, P. & Stadler, P. A. *Nature* **316**, 126-131 (1985).
27. Segal, M. *J. Physiol., Lond.* **303**, 423-439 (1980).
28. Aghajanian, G. K. & Lakoski, J. M. *Brain Res.* **305**, 181-185 (1984).
29. Beck, S. G., Clarke, W. P. & Goldfarb, J. *Eur. J. Pharmac.* **116**, 195-197 (1985).

Induction of cytotoxic T-cell responses *in vivo* in the absence of CD4 helper cells

R. Mark L. Buller*, Kevin L. Holmes†, Ambros Hügin*, Torgny N. Frederickson† & Herbert C. Morse III†

* Laboratory of Viral Diseases and

† Laboratory of Immunopathology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

Cytotoxic T lymphocytes (CTL) seem to provide the major line of defence against many viruses. CTL effector functions are mediated primarily by cells carrying the CD8 (Ly-2) antigen (CD8⁺ cells) and are triggered by interactions of the T-cell receptor with an antigenic complex, often termed 'self plus X', composed of viral determinants in association with class I molecules of the major histocompatibility complex (MHC). The mechanism(s) of induction of virus-specific CTL *in vivo* is poorly understood, but data from *in vitro* experiments suggest that their generation is strictly dependent on functions provided by CD4⁺ helper T cells (also referred to as L3T4⁺; or T_H) that respond to antigens in the context of class II (Ia) MHC determinants¹⁻³. The prevailing opinion that induction of most functions of CD8⁺ cells requires help provided by CD4⁺ cells has recently been challenged by the observation that CD8⁺ cells alone can mediate a variety of responses to alloantigens *in vitro* and *in vivo*⁴⁻⁷; however, the possibility that CTL to self plus X could be generated *in vivo* in the absence of T_H cells has not been evaluated. We report here that C57BL/6J (B6) and AKR/J mice, when functionally depleted of CD4⁺ cells by *in vivo* treatment with the CD4⁺-specific rat monoclonal antibody GK1.5 (refs 8-14) responded to ectromelia virus infection¹⁵⁻¹⁸ by developing an optimal *in vivo* virus-specific CTL response¹⁹⁻²³, and subsequently recovered from the disease (mousepox) that was lethal for similarly infected nude mice (CD4⁻, CD8⁻).

To determine the role of CD4⁺ cells in the development of this CTL response, B6 mice were injected intraperitoneally (i.p.) at specific intervals (legend, Table 1) with monoclonal antibody (mAb)GK1.5. Injected mice were examined at various times over the next four weeks for the frequencies of CD8⁺ and CD4⁺ cell subsets in lymph nodes and/or spleen as determined by

flow microfluorometry (FMF) using fluorescein isothiocyanate (FITC)-labelled mAbs. The results of a representative experiment are given in Table 1. The FMF studies showed that the frequencies of CD4⁺ cells in spleens of mice five days after the last inoculation of mAb GK1.5 (time *t* = 7 days) were comparable to those detected in spleens of nude mice. Similar levels of depletion were observed as soon as two days after the second mAb treatment (*t* = 0 days) and could be maintained for 28 days (the longest time examined) by repeating injections of mAb GK1.5 every six days after the initial three treatments (data not shown). It was possible that the frequencies of CD4⁺ cells detected with FITC-labelled mAb GK1.5 were inappropriately low as a result either of unlabelled mAb GK1.5, used in treatment, blocking the binding of the FITC-labelled mAb GK1.5, or the loss of antigen from the surface of viable T_H cells. These possibilities were considered unlikely because of two observations. First, if the CD4 antigens were masked by unlabelled mAb GK1.5, these cells would be detectable using the FITC-labelled antibody against rat immunoglobulin. Few cells reactive with antibody to rat immunoglobulin were detected (Table 1). Second, the frequencies of CD4⁺ plus CD8⁺ cells in spleens of both normal and treated mice closely approximated the frequencies of Thy-1.2⁺ cells. T_H cells are both CD4⁺ and Thy-1.2⁺. Thus if T_H cells were present in the spleens of treated animals, but devoid of surface CD4⁺ antigen, then the frequencies of Thy-1.2⁺ cells would be expected to exceed the sum of the CD8⁺ and CD4⁺ subsets of T cells.

Two studies were also performed to determine if there was any functional evidence for the persistence of CD4⁺ cells in mice treated with mAb GK1.5. First, normal, nude and mAb GK1.5-treated mice were immunized with the T_H-dependent antigen, sheep red blood cells (SRBC), and the plaque-forming cell (PFC) responses of their spleen cells were tested four days later (Table 1). The immunoglobulin (IgM) PFC responses of mice treated with mAb GK1.5 were markedly reduced compared to those of normal mice and were indistinguishable from the PFC responses generated by nude mice. Second, control and GK1.5-treated mice were infected with ectromelia virus and tested for neutralizing antibody titres at various times after infection (Table 1). Infected mice treated with PBS developed high titres of neutralizing antibodies that peaked at day 21 after infection. By comparison, very low titres of neutralizing antibody were detected in sera of mice treated with mAb GK1.5. These results demonstrated that mice treated with GK1.5 were functionally devoid of the CD4⁺ cells required to generate significant antibody responses to SRBC or ectromelia virus.

To assess the direct effect of CD4⁺ cell depletion on the induction of a CTL response, we investigated the ability of treated B6 mice to elicit an ectromelia virus-specific CTL response. Figure 1 depicts the virus-specific CTL responses of control and CD4-depleted mice at 7 (Fig. 1a, b) and 11 (Fig. 1c, d) days after infection. At neither time were there significant differences between the CTL activities generated by spleen cells of control and treated animals. Results similar to those in Fig. 1a and b were obtained in three other experiments involving six control and six mAb-GK1.5-treated mice (data not shown). For the mice used as sources of spleen cells in these studies, the efficacy of the protocol for depleting CD4⁺ cells was monitored in one of two ways. First, cagemates of the control and treated mice used for the experiments in Fig. 1a and b were immunized with SRBC three weeks after infection with ectromelia virus. The PFC responses of treated mice were comparable to those of the treated animals examined earlier (Table 1). Second, aliquots of the spleen cells used for the CTL assays shown in Fig. 1c and d were examined by FMF. The frequencies of CD4⁺ cells in spleens of treated mice were also similar to those of treated mice reported in Table 1. The development of a normal CTL response by treated mice was thus not due to the re-emergence of CD4⁺ cells in the spleens of infected animals. Note that in these assays, there was no significant lytic activity

Table 1 Immunological characteristics of mice treated with monoclonal antibody against CD4

Mice strain	genotype	Treatment*	Frequencies of antigen-positive cells†				Anti-SRBC response‡		Anti-ectromelia virus neutralizing response (days after infection)§			
			Thy-1 ⁺	CD4 ⁺	CD8 ⁺	rIg ⁺	PFC per 10 ⁶ cells	PFC per spleen	10	14	21	28
B6	+/+	PBS	33 (0.9)	24 (0)	11 (0.9)	<1	649 (188)	42,208 (10,356)	218 $\pm$ 1.4	356 $\pm$ 1.3	4,467 $\pm$ 1.5	2,313 $\pm$ 2.6
B6	+/+	GK1.5	22 (2.0)	3 (0.9)	16 (1.2)	2 (0.6)	25 (9)	1,033 (303)	41 $\pm$ 1.0	41 $\pm$ 1.0	69 $\pm$ 1.3	122 $\pm$ 2.0
B10	+/ <i>nu</i>	None	21 (2.4)	14 (1.8)	7 (1.2)	ND	542 (155)	42,333 (10,424)	ND	ND	ND	ND
B10	<i>nu/nu</i>	None	2 (0.4)	2 (0.3)	1 (0)	ND	55 (23)	3,687 (1,296)	ND	ND	ND	ND

* C57BL/6J (B6) mice were inoculated i.p. with phosphate-buffered saline (PBS) or PBS containing 1 mg of mAb GK1.5 at days -4, -2, 2 and every six days thereafter to deplete CD4⁺ cells. Monoclonal antibody GK1.5 was prepared as a 50% ammonium sulphate precipitate of ascites obtained from NIH *nu/nu* mice injected i.p. with 1–2 $\times$ 10⁷ hybridoma cells. The antibody concentration was determined by a radial immunodiffusion assay using rat IgG standards.

† Mean frequencies of spleen cells from three individual mice reactive with FITC-labelled antibodies to Thy-1.2, CD4, CD8, and rat immunoglobulin (rIg) as detected by FMF^{29,30}. The anti-rat antibody was used at saturating concentrations, previously determined using unlabelled mAb GK1.5 reacted with normal mouse spleen cells. Numbers in parentheses, one s.e.m.

‡ B6 mice were immunized i.p. at day 3 with 0.2 ml of a 10% suspension of SRBC and killed at day 7. C57BL/10ScN (B10) mice were killed four days after immunization with SRBC. Figures indicate mean numbers of direct plaque-forming cells for three individual mice. Numbers in parentheses indicate one s.e.m. B6 and B10 *nu/nu* mice not inoculated with SRBC had a background level of PFC per spleen of ~37 (25) and 56 (43), respectively. PFC assays were performed as described³¹.

§ On day 0, mice were inoculated in the left rear footpad with 0.05 ml of PBS containing 5 $\times$ 10⁴ PFU ectromelia virus (strain 79)¹⁶. At the indicated times post-infection (PI), plasma was obtained from three PBS controls and four mice treated with mAb GK1.5. A neutralization assay was carried out on the plasma, and the titre was expressed as a reciprocal of the dilution which resulted in a 50% reduction in the number of indicator plaques³². Figures indicate the geometric mean and one relative standard error. All PBS treated mice were significantly different from mAb GK1.5 treated mice ($P < 0.05$, Student *t*-test). Titres in sera obtained from these mice before infection were all <40.

Table 2 Mortality in B6 and B10 mice infected with ectromelia virus

Strain	Genotype	Treatment	Virus inoculum (PFU)*		5 $\times$ 10 ²	5 $\times$ 10 ¹	Overall mortality†
			5 $\times$ 10 ⁶	5 $\times$ 10 ⁵			
B6	+/+	PBS			0/10		0/10
B6	+/+	GK1.5‡			1/11		1/11
B10ScN	+/ <i>nu</i>	None	0/3	0/3	0/3	0/3	0/12
B10ScN	<i>nu/nu</i>	None		3/3	3/3	3/3	4/4
							16/16

* Mice were inoculated in the left rear footpad with 0.05 ml PBS containing the indicated quantity of virus.

† Number of deaths over the number of mice inoculated over a 42-day observation period. Mean day of death for *nu/nu* mice ranged from 9.3 at 5 $\times$ 10² PFU to 13.8 at 5 $\times$ 10¹ PFU.

‡ Treatment of mice with mAb GK1.5 was as in Table 1.

against uninfected EL-4 cells or infected or uninfected H-2-incompatible (H-2^d) P815 cells (Fig. 1). These results indicate that target cell lysis was both H-2-restricted and virus-restricted, and could not be ascribed to the activity of natural killer cells.

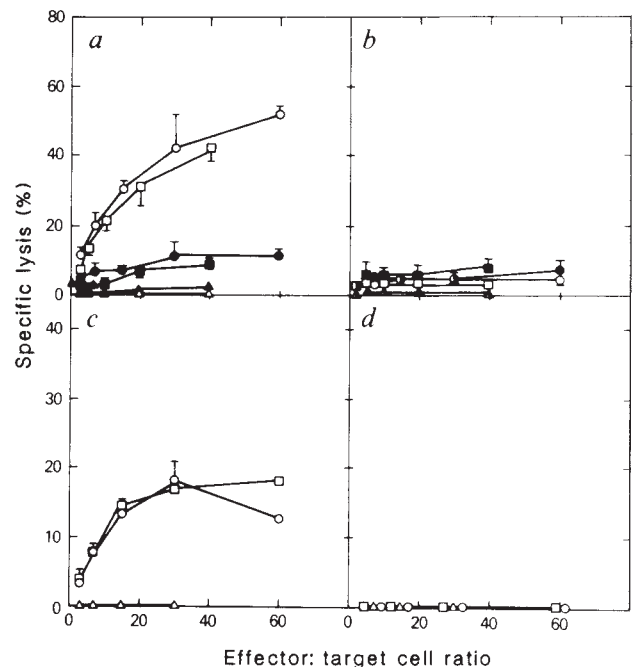
These data demonstrated that normal ectromelia virus-specific CTL responses could be generated *in vivo* in the absence of functional CD4⁺ cells. In view of previous reports correlating recovery from mousepox with the presence of a specific CTL response^{19–23}, it was expected that mice lacking CD4⁺ cells would survive infection with the virus whereas nude mice, deficient in CD8⁺ precursors of CTL, as well as CD4⁺ cells, would succumb to virus challenge. To evaluate this possibility, control and mAb-GK1.5-treated B6 mice and untreated B10 +/*nu* and *nu/nu* mice were infected with ectromelia virus and observed for morbidity, mortality and histological evidence of disease (Table 2). All *nu/nu* mice died, regardless of the virus dose used for challenge, whereas +/*nu* and PBS-treated mice survived the infection with no morbidity. One of 11 B6 mice treated with mAb GK1.5 died ten days after challenge with 5 $\times$ 10⁴ plaque-forming units (PFU) of virus. The cause of death of this animal is not known because severe lesions of the spleen and liver, characteristic of mice dying with ectromelia virus, were not observed at autopsy. Additional histological studies of tissues obtained 7 days after infection showed that necrotic lesions in spleen, liver and lymph nodes of treated mice were more severe than those of control mice; however, by 11 days after infection, the number and size of necrotic lesions were reduced for both sets of mice, and infectious virus was not recovered from spleen cells of either group. The difference in the course of recovery for mice treated with mAb GK1.5 could be due to the absence of T_H-dependent virus-specific antibody response and/or production of γ -interferon, a product of CD4⁺

cells, that together would act to limit virus spread in untreated mice before CTL exert their full effect. These results demonstrated that resistance of B6 mice to lethal infection with ectromelia virus was T-cell dependent and that under these experimental conditions, neither CD4⁺ cells nor an anti-ectromelia virus antibody response (Table 1) was required for this effect.

To determine if this *in vivo* mechanism of induction of resistance to viral infections could be generalized beyond the response of B6 mice to ectromelia virus, two additional studies were undertaken. First, AKR/J mice were depleted of CD4⁺ cells and challenged with ectromelia virus. Six days later, virus-specific CTL responses of spleen cells from three individual antibody-treated and three PBS-treated mice were assayed and shown to be equivalent (22 $\pm$ 4% compared to 26 $\pm$ 2% specific lysis, respectively, at effector/target ratios of 60:1). Second, B6 mice depleted of CD4⁺ cells were challenged with inocula of vaccinia virus (1 $\times$ 10⁶ PFU) lethal for nude mice. Again, CTL responses of spleen cells from mAb-GK1.5-treated mice were equivalent to those of PBS-treated mice when assayed seven days after infection (data not shown). Finally, preliminary studies have shown that B6 mice depleted of CD4⁺ cells are resistant to lethal infection with murine cytomegalovirus (G. M. Shearer and A. Singer, personal communication).

These results imply that CTL responses mediated by CD8⁺ cells are sufficient to resist lethal infection with different viruses. Attempts to demonstrate this point directly by long term depletion *in vivo* of CD8⁺ cells with either of two monoclonal antibodies against CD8 (mouse mAb 19/178 or rat mAb 53-6.7) have not been totally successful. Using a treatment regimen similar to that employed for mAb GK1.5, only transient depletion was obtained with CD8⁺ cells reappearing ~14–20 days

Fig. 1 Primary *in vivo* CTL responses of spleen cells from control and mAb-GK1.5-treated mice infected with ectromelia virus. Female B6 mice were injected on days -4, -2 and 3 with 0.2 ml of PBS (○) or PBS containing 0.5–1.0 mg of mAb GK1.5 (□) or left untreated (Δ). On day 0, the treated mice were inoculated in the footpad with 0.05 ml of PBS containing 5×10^4 PFU of ectromelia virus (strain NIH79). Spleen cell suspensions obtained from mice 7 days (a, b) or 11 days (c, d) after infection were tested for lytic activity in a 4-h ^{51}Cr release assay using targets infected with an antigenically cross-reacting vaccinia virus (a, c) or uninfected target cells (b, d). The ^{51}Cr -labelled targets were H-2-compatible (H-2^b) EL-4 cells (open symbols) and H-2-incompatible (H-2^d) P815 cells (closed symbols). Points indicate per cent specific lysis for individual mice (no standard error bars) or the mean value for two or three individual mice with standard error bars indicating two s.e.m. Spontaneous ^{51}Cr release was always <27% of maximum release.



following the initiation of treatment, possibly as a result of an anti-idiotypic or anti-rat antibody response by the mouse to the mAb administered. In spite of these limitations, it was found that one of five B6 mice treated with mAb 19/178 and challenged with ectromelia died of severe virus-induced disease 13 days after inoculation, and that spleens from two mice killed at day 15 contained more than 10^2 PFU per 0.1 g, whereas spleens from control mice were virus negative. Similar results were obtained with mAb 53-6.7, except that no deaths were observed. These observations underscore the importance of CD8⁺ cells in recovery from ectromelia virus infection.

These results are consistent with at least three models for the *in vivo* generation of CTL specific for viruses in mice. In the first model, induction of CTL to virus does not require help provided by CD4⁺ cells and is dependent solely on the response of CD8⁺ cells to self plus X. The response of CD8⁺ cells may or may not include the need to produce and respond to interleukin-2 or other lymphokines as suggested by earlier reports^{24–27}. The possibility that, in the present study, a residual population of T_H cells expressing very low levels of CD4 antigen may be sufficient to augment the proliferation of virus-specific CD8⁺ cells cannot be excluded. Second, there may be two distinct lineages of CD8⁺ CTL precursors, one dependent on, and the other independent of, T_H for their maturation. The T_H-independent population may provide the predominant response to at least some viruses, resulting in no apparent loss of CTL function in the absence of CD4⁺ cells^{4–7}. In a third model, the CTL response to many antigens may progress from an early T_H-independent stage to a later one of T_H-dependence^{6,7,26}. The response required for resistance to ectromelia or cytomegaloviruses may be too short for a T_H-dependent phase of CTL function to be observed.

It will be important to assess if the results presented here can be generalized to include CD4⁺-cell independent *in vivo* CTL responses to additional pathogens by strains other than B6 and AKR. Earlier studies showed that the *in vitro* proliferative response of purified CD8⁺ cells to a class-I alloantigen stimulus is highly strain-dependent with cells from B6 or B10 mice being the most responsive⁷. One interpretation of these data, consistent with the second model proposed above, is that the balance between T_H-dependent and T_H-independent CTL responses to a specific antigen is genetically determined. If this is true, then CTL responses of other mouse strains to ectromelia virus may

be biased towards a pathway requiring T_H cells. It has also been found that, for a single mouse strain, there is a hierarchy of responses by CD8⁺ cells to different antigens that cannot be explained by variations in precursor frequencies²⁸. If our results reflect antigen-specific differences in the proportions of responding CD8⁺ cells that require or are independent of T_H, the *in vivo* CTL response of B6 mice to pathogens other than ectromelia, vaccinia or murine cytomegalovirus could prove to be completely T_H-dependent or to include both T_H-dependent and -independent responses.

We thank Dr Philip Baker for reagents and advice; Mr Cornelio Duarte for technical assistance; Dr A. Kruisbeek for mAb 19/178; Drs Wendy Davidson, Alfred Singer, Jonathan Sprent and Karen Elkins for critical reviews of the manuscript; Ms Brenda Marshall for secretarial assistance. R.M.L.B. thanks Dr Bernard Moss for continued support.

Received 27 February; accepted 14 April 1987.

1. Cantor, H. & Boyse, E. A. *J. exp. Med.* **141**, 1390–1399 (1975).
2. Bach, F. H. *et al. Immun. Rev.* **35**, 76–96 (1977).
3. Wagner, H. & Rollinghoff, M. *J. exp. Med.* **148**, 1523–1538 (1978).
4. Singer, A., Kruisbeek, A. M. & Andrysiak, P. M. *J. Immun.* **132**, 2199–2209 (1984).
5. Guimezanes, A. & Schmitt-Verhulst, A.-M. *Eur. J. Immun.* **15**, 1187–1191 (1985).
6. Sprent, J. & Schaefer, M. *J. exp. Med.* **162**, 2068–2088 (1985).
7. Sprent, J., Schaefer, M., Lo, D. & Korngold, R. *J. exp. Med.* **163**, 998–1011 (1986).
8. Dialynas, D. P. *et al. J. Immun.* **131**, 2445–2451 (1983).
9. Dialynas, D. P. *et al. Immun. Rev.* **74**, 29–56 (1983).
10. Wofsy, D., Mayes, D. C., Woodcock, J. & Seaman, W. E. *J. Immun.* **135**, 1698–1701 (1985).
11. Coulie, P. G., Coutelier, J.-P., Uytendaele, C., Lambotte, P. & Van Snick, J. *Eur. J. Immun.* **15**, 638–640 (1985).
12. Titus, R. G., Ceredig, R., Cerottini, J.-C. & Louis, J. A. *J. Immun.* **135**, 2108–2114 (1985).
13. Waldor, M. K. *et al. Science* **227**, 415–417 (1985).
14. Wofsy, D. & Seaman, W. E. *J. exp. Med.* **161**, 378–391 (1985).
15. Fenner, F. *Aust. J. exp. Biol. med. Sci.* **25**, 275–282 (1947).
16. Allen, A. M., Clarke, G. L., Ganaway, J. R., Lock, A. & Werner, R. M. *Lab. Anim. Sci.* **31**, 599–608 (1981).
17. Wallace, G. D. & Buller, R. M. L. *Lab. Anim. Sci.* **35**, 41–46 (1985).
18. O'Neill, H. C. & Blanden, R. V. *Infect. Immun.* **41**, 1391–1394 (1983).
19. Blanden, R. V. *J. exp. Med.* **132**, 1035–1054 (1970).
20. Blanden, R. V. *J. exp. Med.* **133**, 1074–1089 (1971).
21. Gardner, I., Bowern, N. A. & Blanden, R. V. *Eur. J. Immun.* **4**, 63–67 (1974).
22. Gardner, I., Bowern, N. A. & Blanden, R. V. *Eur. J. Immun.* **4**, 68–72 (1974).
23. O'Neill, H. C., Blanden, R. V. & O'Neill, T. J. *Immunogenetics* **18**, 255–265 (1983).
24. Widmer, M. B. & Bach, F. H. *Nature* **294**, 750–752 (1981).
25. Von Boehmer, H. & Turton, K. *Eur. J. Immun.* **13**, 176–179 (1983).
26. Von Boehmer, H., Kisielow, S., Leiserson, W. & Haas, W. *J. Immun.* **133**, 59–64 (1984).
27. Moldwin, R. L., Lancki, D. W., Herold, K. C. & Fitch, F. W. *J. exp. Med.* **163**, 1566–1582 (1986).
28. Sprent, J., Schaefer, M., Lo, D. & Korngold, R. *Immun. Rev.* **91**, 195–218 (1986).
29. Morse, H. C. III *et al. J. Immun.* **122**, 443–454 (1979).
30. Yeh, C.-J. G., Hsi, B.-L. & Faulk, W. P. *J. Immun. Meth.* **43**, 269–275 (1981).
31. Mishell, R. I. & Dutton, R. W. *J. exp. Med.* **126**, 423–442 (1967).
32. Buller, R. M. L. & Wallace, G. D. *Lab. Anim. Sci.* **35**, 473–477 (1985).

Translocation of a localized maternal mRNA to the vegetal pole of *Xenopus* oocytes

D. A. Melton

Department of Biochemistry and Molecular Biology,
Harvard University, 7 Divinity Avenue, Cambridge,
Massachusetts 02138, USA

A prominent hypothesis in embryology is that localized maternal factors are important in specifying cell fate. There are, however, only a few examples of maternal molecules that have been shown to be localized and very little is known about how such factors are physically localized within an egg (for review see ref. 1). Previously, cDNA clones were obtained for a class of localized maternal mRNAs from *Xenopus laevis*². These mRNAs are unusual in that they are concentrated at either the animal or vegetal pole of unfertilized eggs. In the present study the synthesis and intracellular distribution of one of them, Vg1, has been examined during oogenesis. The results show that Vg1 mRNA is localized as a crescent at the vegetal pole of mature oocytes. Surprisingly, this mRNA is uniformly distributed in the cytoplasm of immature oocytes. These findings suggest that a single cell, the frog oocyte, has some mechanism for translocating specific RNAs like Vg1. The process that moves Vg1 mRNA is evidently a cytoplasmic localization machinery which is not directly coupled to the synthesis of Vg1 RNA.

In a previous study, total RNA was isolated from animal or vegetal pole caps cut from frozen unfertilized eggs. Northern blots of these RNA preparations showed that Vg1 mRNA is 20–25 times more abundant in the vegetal pole or presumptive endoderm of the egg². To determine more precisely where Vg1 RNA is located and to investigate how it comes to be there, *in situ* hybridizations were performed on sectioned oocytes. Oocytes are the ovarian cells which undergo a steroid-induced maturation into eggs. The results show that Vg1 RNA is located in a cortical crescent in the vegetal hemisphere (Fig. 1). This thin (~10 µm) crescent begins just beneath the equator and forms a uniform bowl at the vegetal end. Hybridization to serial sections shows that the crescent of Vg1 RNA localization is seen throughout the radially symmetric oocyte (Fig. 1c) and not only below the nucleus. It is noteworthy that the Vg1 RNA is restricted to a thin subcortical shell and does not spread from the vegetal pole in a gradient. These *in situ* hybridizations strongly suggest that the vegetal placement of Vg1 RNA in unfertilized eggs is a direct consequence of its location in fully grown oocytes.

Several kinds of assays show that most oocyte poly(A)⁺ RNAs are not localized and are uniformly distributed in the cytoplasm. Differential screening of cDNA libraries with probes specific for animal and vegetal pole RNAs and *in vitro* translation of messenger RNAs isolated from animal and vegetal sections both support this conclusion^{2,3}. The typical and even RNA distribution of maternal mRNAs is exemplified by histone H4 mRNA. Histone H4 mRNA, which is at least 20 times more abundant than Vg1 RNA, is found throughout the cytoplasm (Fig. 1d) as was previously shown by Jamrich *et al.*⁴. Shorter exposures show that there is a smooth gradient of histone mRNA from the animal to vegetal pole and that the animal pole has about 2–3 times more H4 mRNA. The shallow animal to vegetal gradient of total poly(A)⁺ RNA may simply reflect the distribution of non-yolk cytoplasm in the oocyte^{5,6}.

To learn how Vg1 RNA is localized it is important to know when during oogenesis the RNA is made and whether it is immediately concentrated at the vegetal pole. If this maternal RNA were always found in a crescent at the vegetal pole of the oocyte, that might suggest that the localization of the RNA is coupled to its synthesis and transport from the nucleus. The

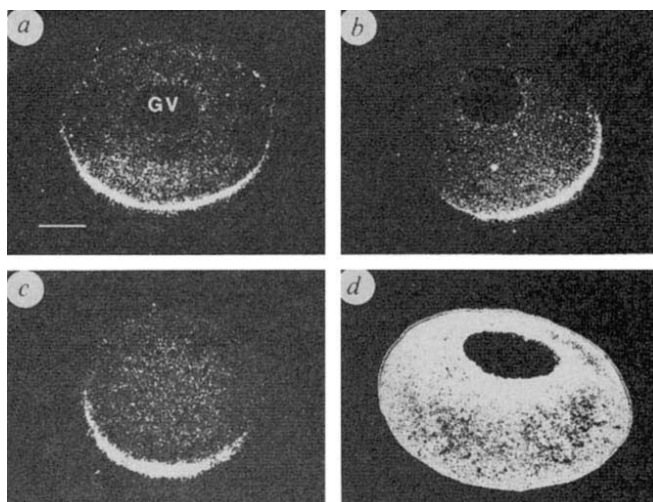


Fig. 1 Localization of Vg1 RNA at the vegetal pole of *Xenopus* oocytes by *in situ* hybridization. Post-vitellogenic (stage V–VI) albino oocytes were fixed in ethanol acetic acid chromium trioxide, dehydrated, embedded in Paraplast and sectioned at 6 µm. Oocytes from albino frogs were used because the pigment granules that mark the animal hemisphere of wild-type oocytes are almost indistinguishable from autoradiograph grains. Following attachment to silane-treated slides, the oocyte sections were treated with proteinase K and prepared for hybridization with single-stranded ³²P-RNA probes (3×10^6 d.p.m. µg⁻¹; 0.3 ng µl⁻¹) as described elsewhere²⁰. After washing and digestion with RNase A, the slides were dipped in Kodak NTB-2 emulsion, exposed for 3–7 days, developed, stained with Giemsa and photographed under darkfield optics. In each photo the animal pole is at or near the top. The nucleus or germinal vesicle (GV) is seen as a circle nearer the animal pole, except in c where the section is taken at a position past the nucleus. The hybridization probe for a–c was ³²P-labelled antisense Vg1 RNA of 2.3 kilobases (kb), and for d was ³²P-labelled antisense histone H4 RNA (0.4 kb). Controls with sense Vg1 probes do not show any autoradiograph signal above background, nor do sections that are treated with RNase before hybridization with Vg1 antisense RNA probes. All photographs were taken at the same magnification and the scale bar in a is 200 µm.

time during oogenesis at which Vg1 RNA is synthesized was assayed by RNase protection using total RNA extracted from oocytes ranging from small pre-vitellogenic to fully grown, post-vitellogenic oocytes. The results (Fig. 2) show that Vg1 RNA is present in the smallest oocytes tested (stage I) and that the level of Vg1 RNA remains constant throughout oogenesis. The assay for Vg1 RNA levels was performed five separate times on oocytes from five different females and together these data allow one to conclude that the total amount of Vg1 RNA does not change by more than a factor of two from stage-II to stage-VI oocytes. Although a continual synthesis and degradation of the RNA cannot be ruled out, the simplest interpretation is that all the Vg1 RNA is made early (by stage II) in pre-vitellogenic oocytes and this store of maternal RNA remains in the growing oocyte and is inherited by the egg.

The location of Vg1 RNA in young oocytes is revealed by the *in situ* hybridizations in Fig. 3. The results show that Vg1 RNA is uniformly distributed in the cytoplasm of young pre-vitellogenic (stage-I and stage-II) oocytes. These young oocytes cannot be oriented for sectioning because there is no morphological sign of the future animal or vegetal pole. However, in no case out of more than 50 oocytes tested was any localization of the autoradiographic signal observed. Thus, when the Vg1 RNA is first synthesized, early in oogenesis, the RNA is evenly spread throughout the cytoplasm and shows no sign of localization.

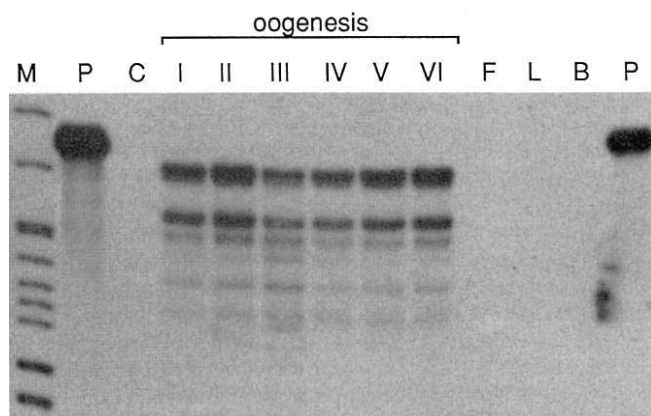


Fig. 2 Vg1 RNA is synthesized in young pre-vitellogenic oocytes. Total RNA was extracted from various stages (stages I–VI as in ref. 21) of oocyte development, pre-vitellogenic to fully grown post-vitellogenic oocytes. Because of the difference in the amount of yolk present at these different stages, RNA recovery was monitored by mixing in ^{32}P -labelled synthetic globin RNA with the oocyte homogenates. The ^{32}P label recovered was used to normalize the amount of total RNA used for the RNase protection assay. RNase protection assays²² were performed with one oocyte-equivalent of RNA. Total RNA from the follicles (F) surrounding 100 stage-IV to stage-VI oocytes and a comparable amount of total RNA from liver (L) or blood (B) were tested. Each RNA sample was hybridized to a uniformly ^{32}P -labelled antisense Vg1 RNA synthesized with SP6 RNA polymerase from *Dde*I-cut pSP65-Vg1 DNA². Following hybridization and digestion with RNase A and RNase T1, the products were fractionated in an 8 M urea 6% polyacrylamide gel. Lane P is full-length probe (370 nucleotides) carried through the entire hybridization procedure, but not digested with RNases. Lane C is the probe digested with RNases following hybridization to tRNA. The antisense probe is protected from RNase digestion by Vg1 transcripts to give two major bands as shown previously². Markers (M) are *Hpa*II-cut pBR322 DNA.

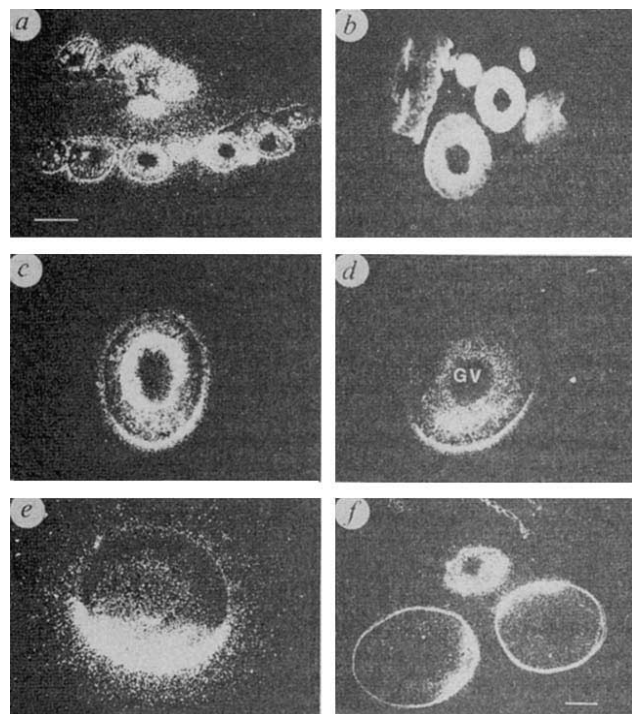


Fig. 3 Translocation of Vg1 RNA during oogenesis. *In situ* hybridizations were performed with an antisense Vg1 RNA probe as described in Fig. 1 legend. Albino oocytes at various stages of growth are shown. The black hole present in the centre of most sections is the nucleus or germinal vesicle (GV). The stages according to Dumont²¹ are I, II, III, III and IV for a–e, respectively. The scale bar in a is 200 μm and applies to a–e. A stage-II and two stage-IV oocytes from a wild-type female are shown in f, where the scale bar is 200 μm . Note that the white ring around the pigmented oocytes is due to the pigment granules in the cortex.

As oocytes grow they take up vitellogenin by pinocytosis from the blood and store derivatives of this protein in yolk platelets⁷. It is during this period (vitellogenesis) that the Vg1 RNA becomes localized. Figure 3c–e shows the beginning of Vg1 RNA localization. The first sign of localization is the concentration of Vg1 RNA near the cortex at one end of the oocyte. In some cases (Fig. 3c) the RNA seems to concentrate around the nucleus as well. In other cases, however, this perinuclear concentration is not observed (Fig. 3d and e). In contrast to Vg1 RNA and as observed in fully grown oocytes (Fig. 1) histone H4 mRNA is uniformly distributed at all these early oocyte stages. These data with histone H4 probes (not shown) confirm studies⁴ which demonstrated the even cytoplasmic distribution of histone RNA during early oogenesis.

The albino oocytes used for the *in situ* hybridizations in Fig. 3a–e lack cortical pigment granules. In wild-type oocytes, cortical pigment granules first appear over the entire surface (at stage III) and are later concentrated at the animal hemisphere (stage IV). Thus, in addition to cell diameter, the appearance of cortical pigment helps mark the stage of oocyte development. The hybridizations in Fig. 3f were performed with pigmented (wild-type) oocytes and these sections highlight the difference in Vg1 distribution between early and later stages in oogenesis. By the time cortical pigment granules have appeared (the white ring on the perimeter of larger oocytes in Fig. 3f) the Vg1 RNA signal is localized at one end.

At a superficial level some aspects of Vg1 RNA synthesis and localization parallel the biogenesis of yolk platelets⁸. Both are initially found throughout the cytoplasm and later form a gradient along the animal-vegetal axis. There are however,

several important differences. Vg1 RNA is synthesized before vitellogenesis begins and the localization of Vg1 RNA is nearly complete at a time when the yolk platelets are still of a uniform size and radially distributed in the cytoplasm (see Fig. 3e and ref. 8). In addition, the final gradient of yolk in the oocyte is not nearly as steep as the quantitative localization of Vg1 RNA. Nevertheless, it is not unreasonable to ask whether Vg1 RNA is imported into oocytes as is vitellogenin. A sensitive test for Vg1 RNA in the follicle, liver and blood (the latter tissues being sites for vitellogenin synthesis and transport, respectively) proved negative (Fig. 2). A layer of follicle cells envelopes every oocyte until maturation at which time the oocyte is released from the follicular layers. In these assays (Fig. 2) RNA from the follicle cells surrounding 100 oocytes and an equal amount of RNA from liver or blood is compared to the RNA found in just one oocyte. The results clearly show that there is at least 100 times more Vg1 RNA in the oocyte than in the other tissues tested. Although the possibility that follicle cells or other tissues synthesize Vg1 RNA and quickly transport it into the oocyte cannot be absolutely ruled out, the results strongly suggest that Vg1 RNA is synthesized by the oocyte itself.

The results of the *in situ* hybridizations (Fig. 3) show that the Vg1 RNA is first found throughout the cytoplasm and is subsequently localized at one end. To test whether this change in location corresponds to any substantial change in the structure of Vg1 RNA, Northern blots containing RNAs isolated from oocytes before, during, and after the localization process were probed for Vg1 sequences. The results (Fig. 4) show that Vg1 RNA is in the form of a mature (spliced) polyadenylated message at all oocyte stages. There is a trace of unadenylated

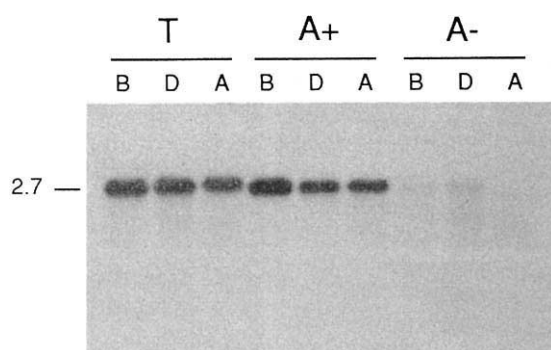


Fig. 4 Mature Vg1 mRNA is present at all stages of oogenesis. RNA was extracted from oocytes before (B), during (D), and after (A) the localization of Vg1 RNA. Using the *in situ* hybridizations shown in Figs 1 and 3 as a guide, this required RNA from stage I–II, stage III–IV and stage V–VI oocytes, respectively. Total RNA (T) was fractionated into poly(A)⁺ and poly(A)[−] fractions by oligo-(dT)-cellulose chromatography. The RNA samples were electrophoresed in formaldehyde-agarose gel, blotted to GeneScreen, and hybridized with ³²P-labelled antisense Vg1 RNA as described previously². Unpublished data, including the DNA sequence of a Vg1 cDNAs, shows that the gene that encodes Vg1 is interrupted by introns and that the mature spliced Vg1 mRNA is about 2.7 kb long.

Vg1 RNA at all stages, though this could be a result of inefficient binding to oligo(dT)-cellulose. It can therefore be concluded that the localization of Vg1 mRNA does not involve a significant change in the size or structure of Vg1 transcripts.

The data presented here suggest two possible mechanisms for the localization of Vg1 RNA. One is that Vg1 RNA is specifically degraded (by a localized RNase activity) in the animal pole and most of the vegetal end of the oocyte. However, because the total amount of Vg1 RNA levels remains relatively constant during oogenesis (Fig. 2) such a localization mechanism would require a coordinated synthesis and degradation of Vg1 transcripts in different regions of late-stage oocytes. A more likely explanation is that all the Vg1 RNA is translocated by an active process into a crescent at the vegetal end of the oocyte. Although the rate of this localization process cannot be measured from the *in situ* hybridizations, the results suggest that the movement of the Vg1 RNA takes place over many days, perhaps even a few weeks. The growth of a stage-II to a stage-IV oocyte, the period during which Vg1 RNA is localized, can take from two to four weeks *in vitro*⁹.

With respect to the function of the Vg1 RNA and the protein it encodes, the data show that Vg1 mRNA is not distributed like the germinal granules associated with germ cell formation. Germinal granules are closely associated with the mitochondrial cloud and do not extend up to the oocyte equator as does Vg1 RNA (ref. 10). It remains to be determined when Vg1 RNA is translated and whether its protein product is also localized. Nonetheless, the extent of the subcortical distribution of Vg1 RNA (Fig. 1) suggests that the Vg1 protein may be involved in a developmental process common to all vegetal cells of the developing embryo. Two obvious possibilities are that the Vg1 protein is involved in the specification of vegetal cells as endoderm or it may be involved in the induction of mesoderm by endoderm¹¹. Alternatively, a regional activation of the Vg1 gene product could be involved in polarizing the dorsal-ventral axis¹².

It is not known how the animal-vegetal axis is established in an oocyte and the mechanism by which Vg1 RNA is localized remains a puzzle. Wylie and his colleagues¹⁰ noted that the mitochondrial cloud lies on what will become the vegetal pole side of the nucleus in pre-vitellogenic oocytes and the alignment

of a centriole pair and the nucleus¹³ might also mark the beginnings of the animal-vegetal axis. It is possible that the relative positions of the nucleus and mitochondrial cloud set up the animal-vegetal axis and other cellular components or processes might be aligned to this initial polarity. For example, cytoskeletal elements polymerized along the animal-vegetal axis might have an intrinsic polarity which is in turn used to localize Vg1 RNA, yolk platelets⁶, and so on. In any case, it is reasonable to assume that the translocation of Vg1 RNA involves interaction with cytoplasmic proteins and it may be possible to identify these components of the localization machinery by using Vg1 RNA as a probe. The oocyte localization process itself warrants further investigation because it may shed light on how other cells localize RNAs^{14–19} and how eggs localize developmental signals.

This work was supported by a grant from the NIH and by my colleagues R. Harvey, C. Kintner, H. O'Keefe, M. Rebagliati and D. Weeks. I thank M. Danilchik and J. Gerhart for communicating unpublished results and B. Tompkins for albino frogs.

Received 2 March; accepted 13 May 1987.

- Davidson, E. *Gene Activity in Early Development* (Academic, New York, 1986).
- Rebagliati, M., Weeks, D., Harvey, R. & Melton, D. *Cell* **42**, 769–777 (1985).
- King, M. & Barklis, E. *Dev Biol.* **112**, 203–212 (1985).
- Jamrich, M., Mahon, J., Gavis, E. & Gall, J. *EMBO J.* **3**, 1939–1943 (1984).
- Phillips, C. *Dev Biol.* **109**, 299–310 (1985).
- Gerhart, J. in *Biological Regulation and Development* (ed. Goldberger, R. F.) 133–293 (Plenum, New York, 1980).
- Wallace, R. in *Oogenesis* (eds Biggers, J. D. & Schuetz, A.) (University Park Press, Baltimore, 1972).
- Danilchik, M. & Gerhart, J. *Dev Biol.* (in the press).
- Wallace, R. & Misulovin, Z. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5534–5538 (1978).
- Heasman, J., Quarmby, J. & Wylie, C. C. *Dev Biol.* **105**, 458–469 (1984).
- Nieuwkoop, P. *Wilhelm Roux Arch. der. Biol.* **162**, 341–373 (1969).
- Vincent, J. P., Oster, G. & Gerhart, J. *Dev Biol.* **113**, 484–500 (1986).
- Coggins, L. *J. Cell. Sci.* **12**, 71–93 (1973).
- Lawrence, J. & Singer, R. *Cell* **45**, 407–415 (1986).
- Jeffrey, W., Tomlinson, C. & Brodeur, R. *Dev Biol.* **99**, 408–417 (1983).
- Kalthoff, K. & Elbetheiha, A. *J. Embryol. exp. Morph.* **97**, suppl. 181–196 (1986).
- Anderson, K. & Nusslein-Volhard, C. *Nature* **311**, 223–227 (1984).
- Macdonald, P. & Struhl, G. *Nature* **324**, 537–545 (1986).
- Mlodzik, M. & Gehring, W. *Cell* **48**, 465–478 (1987).
- Kintner, C. & Melton, D. *Development* **99**, 311–325 (1987).
- Dumont, J. *J. Morph.* **136**, 153–180 (1972).
- Melton, D. *et al. Nucleic Acids. Res.* **12**, 7035–7056 (1984).

The human tumour-associated epithelial mucins are coded by an expressed hypervariable gene locus *PUM*

Dallas M. Swallow*, Sandra Gendler†, Beatrice Griffiths*, Gerald Corney*, Joyce Taylor-Papadimitriou† & Michael E. Bramwell‡

* MRC Human Biochemical Genetics Unit, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK

† Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK

‡ Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

A single highly-polymorphic autosomal gene locus *PUM* codes for a family of mucin-type glycoproteins, separable by SDS-gel electrophoresis, which we first identified in human urine^{1,2}. The locus also codes for glycoproteins which are abundant in several other normal epithelial tissues and body fluids, including milk, and in tumours of epithelial origin³. These mucin-type glycoproteins seem to be very immunogenic in rodents and, in a search for epithelial specific or tumour-associated antigens, a large number of related antibodies have been isolated^{4–9} which bind to the *PUM*-coded mucins³. Many of the antibodies show a pronounced tumour specificity on immunohistology^{10–13} and are being used

Fig. 1 Detection of the PUM polymorphism. The glycoproteins were separated by SDS 5–15% polyacrylamide gel electrophoresis (PAGE), transferred electrophoretically onto nitrocellulose, and detected using the monoclonal antibody Ca1, as described previously². *a*, Mendelian inheritance of the PUM alleles. Part of family MRC 5188 is shown. Urine samples from the two parents, PUM phenotypes AC and DF were applied in adjacent tracks to samples from four of the children (phenotypes AF, CD, AD and CF) for ease of comparison. *b*, A selection of samples from the two families MRC 5188 and MRC 1489 used for the present study to show the seven alleles which segregate on these families (A–G). (Note that the PUM alleles have been renamed since our first report¹, because of the greater resolution of this technique).

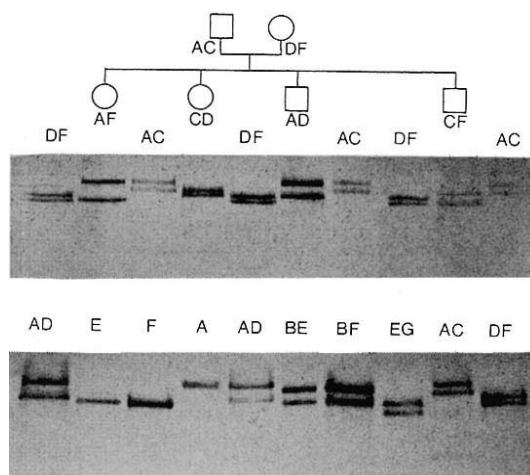
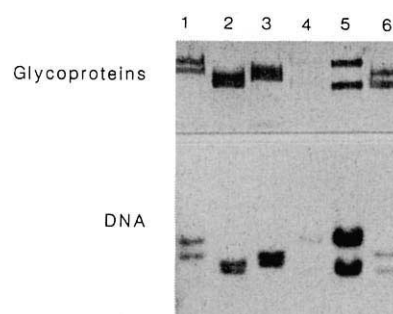


Fig. 2 Comparison of the PUM protein and DNA phenotypes. Part of family MRC 5188 is shown. Lane 1, father, phenotype AC; lane 2, mother, phenotype DF; lanes 3–6, four of the children (two are the same individuals shown in Fig. 1), phenotypes CD, AF, AF and CF. The PUM glycoproteins were detected with Ca1 as described in Fig. 1. The DNA was prepared from white cells³⁰ and digested with *Eco*RI, separated by electrophoresis in 0.7% agarose gels at 1.1 V cm⁻¹ for 18 h, transferred passively onto Biotodyne membrane and the blots probed with pMUC10 (ref. 16).



widely in cancer diagnosis *in vitro*^{9–12} and *in vivo*^{11,14} and even in cancer therapy¹⁵. To investigate the expression of these antigens in normal and malignant cells complementary DNA coding for the mammary mucin has been isolated¹⁶. Here we present evidence obtained using this cDNA that the *PUM* locus is a hypervariable 'minisatellite' region of human DNA similar to those described by several groups^{17–25}, but which is novel in that it is transcribed and translated, and that the same polymorphism is demonstrable in the expressed gene product.

Although the *PUM* (from peanut lectin binding urinary mucins) polymorphism can be detected in human urine by immunoblot analysis using a range of antibodies, the simplest pattern of bands is detected by the monoclonal antibody Ca1 (ref. 2). Each individual shows either one or two bands which behave as codominant allele products (Fig. 1), and at least ten alleles can be easily resolved. Because the codominance of the *PUM* phenotypes had indicated to us that the variation at the *PUM* gene locus was in the protein core of the *PUM* molecule¹, any restriction fragment length polymorphism (RFLP) that we might detect should show complete genetic linkage to the *PUM* locus. We have therefore collected blood samples and prepared DNA from families whose *PUM* type had previously been ascertained by analysis of urine.

Genomic DNA was prepared from members of two large families which comprise six different matings and segregate seven distinct alleles at the *PUM* locus. Electrophoretic analysis of *Eco*RI-digested DNA probed with the recombinant plasmid pMUC10 (ref. 16) revealed extensive person-to-person variation. A very simple pattern of either one or two major bands was found in each individual. Bands differed in size in the range 7–12 kilobases (kb). Family studies showed that the bands represented alleles of a single gene and, because only one band was detected per allele, it was deduced that all the *Eco*RI cutting sites were outside the main region covered by the probe. However, the most remarkable feature of this variation was its

relationship to the protein polymorphism. The alleles detected by pMUC10 not only cosegregated with the *PUM* alleles, showing no recombinants amongst the 20 offspring of the six matings, but surprisingly the relative mobility of all the seven allelic fragments detected by pMUC10 was the same as the relative mobility of the protein allele products detected by Ca1 (Fig. 2).

It thus now seems likely that the differences in the allele products are due to variation in size of the protein core, resulting from person-to-person variation in length of the *PUM* gene. This suggests that there are multiple tandem repeats in the *PUM* gene, allelic variation being due to variation in the number of repeats caused by unequal crossing-over events in the evolutionary past, such as occur in the hypervariable minisatellite regions of DNA^{17–25}. Nucleotide sequence analysis has shown that pMUC10 does indeed contain a repeat of 60 base pairs (S.G., unpublished observations).

In an attempt to improve the resolution of the alleles, by cutting nearer the repeat unit, we have used the enzyme *Hinf*I (refs 21, 22 and 24), which recognizes a 4-bp sequence and cuts frequently in human DNA. The fragment patterns that result are similar to those obtained after digestion with *Eco*RI except that all the bands are smaller (3–7 kb) (Fig. 3). Furthermore, additional alleles can be resolved whose existence was only suspected at the level of the protein and with *Eco*RI-digested DNA (Fig. 3b). Thus the total number of alleles is likely to be very large and the observations so far suggest that the locus may be almost as variable as that described by Wong *et al.*²⁴. The difference in size between the smallest and largest *PUM* fragments is about 3,400 bp, suggesting a maximal difference in the relative molecular mass (M_r) of the polypeptide of ~110,000. In view of the protective role of mucins on the surface of epithelial cells, the functional and biological significance of this extensive variation in *PUM* protein length is intriguing.

None of the hypervariable minisatellite regions of human DNA described^{17–25} have so far been shown to code for protein.

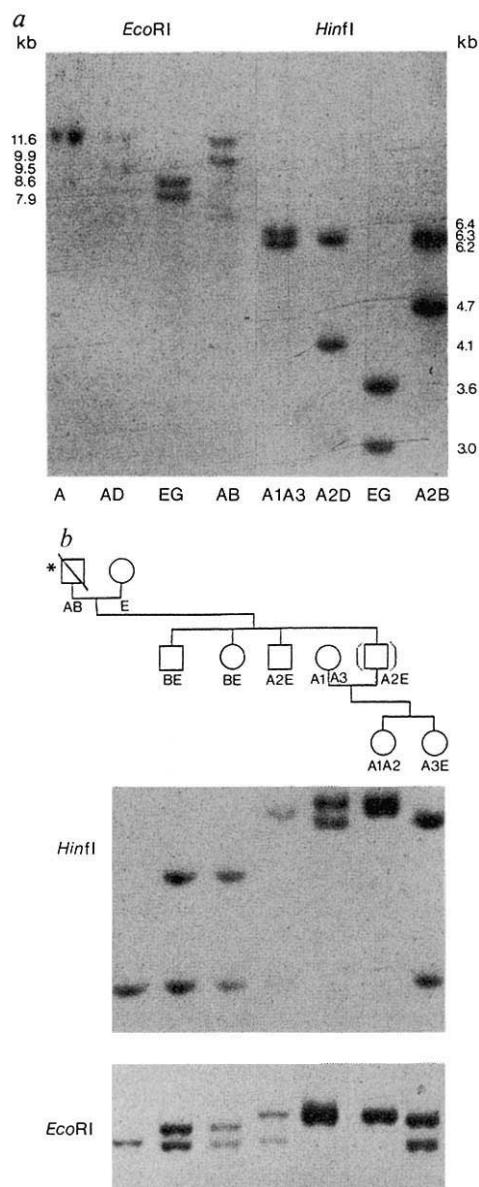


Fig. 3 Comparison of the PUM DNA phenotypes detected using *HinfI*-digested and *EcoRI*-digested DNA and the resolution of the A allele into A1, A2 and A3 by the greater separation of the smaller fragments generated by *HinfI*. *a*, Two parts of the same gel showing the relative mobilities and sizes of the allelic fragments produced by the two restriction enzymes detected with pMUC10. The approximate sizes of the fragments were estimated by comparison with the mobilities of the bacteriophage λ *HindIII* digestion fragments. Electrophoresis was conducted in 0.6% agarose gels at 1.4 V cm^{-1} for 25 h. *b*, Segregation of A1, A2 and A3 alleles in part of family MRC 1489. Photographs of the two gels are placed such that the tracks lie underneath the correct individuals on the pedigree. The upper gel shows the *HinfI* fragments, and the lower gel the *EcoRI* fragments, detected by pMUC10. Note that the A bands show slight variation with *EcoRI* which is resolved much more clearly with *HinfI*. * The urine only of this individual (deceased) was tested, and showed phenotype AB. The presumed phenotype of this individual with *HinfI*-digested DNA is A2B. Parentheses, individual of phenotype A2E whose DNA is not shown on this gel.

However, there are a number of examples of proteins, namely several surface antigens of *Plasmodium*²⁶, silk fibroin²⁷, *Drosophila* glue protein²⁸ and the salivary proline-rich proteins²⁹, which contain multiple small repeats of sequence (4–21 amino acids) and for which allelic variations in length have been described. The common characteristic of these proteins is

that they are rich in particular amino acids. Mucins are characterized by their high serine, threonine and proline content and it seems probable that other mucin genes, like *PUM*, will turn out to have a repetitive structure within their coding regions and may also be hypervariable loci.

We thank Dr D. A. Hopkinson for helpful discussion; MRC families 1489 and 5188 for their kind cooperation and Dr Mary Davis for some of the samples. M.E.B. is the James Hanson Fellow of the Cancer Research Campaign.

Received 26 March; accepted 18 May 1987.

- Karlsson, S. *et al.* *Ann. hum. Genet.* **47**, 263–269 (1983).
- Swallow, D. M., Griffiths, B., Bramwell, M., Wiseman, G. & Burchell, J. *Disease Markers* **4**, 247–254 (1986).
- Griffiths, B. *et al.* *Int. J. Cancer* (submitted).
- Ashall, F., Bramwell, M. E. & Harris, H. *Lancet* **ii**, 1–6 (1982).
- Bramwell, M. E. *et al.* *Cancer* **56**, 105–110 (1985).
- Taylor-Papadimitriou, J. *et al.* *Int. J. Cancer* **28**, 17–21 (1981).
- Ellis, I. O. *et al.* *Histopathology* **8**, 501–506 (1984).
- Foster, C. S., Edwards, P. A. W., Dinsdale, E. A. & Neville, A. M. *Virchows path. Anat. Physiol.* **394**, 279–293 (1982).
- Hilkens, J. *et al.* *Int. J. Cancer* **34**, 197–206 (1984).
- McGee, J. O'D., Woods, J. C., Ashall, F., Bramwell, M. E. & Harris, H. *Lancet* **ii**, 7–10 (1982).
- Epenetos, A. A. *et al.* *Lancet* **ii**, 999–1004 (1982).
- Wilkinson, M. J. *et al.* *Int. J. Cancer* **33**, 299–304 (1984).
- Burchell, J. *et al.* *Cancer Res.* (in the press).
- Pateisky, N. *et al.* *J. nucl. Med.* **26**, 1369–1376 (1985).
- Pectasides, D. *et al.* *Br. J. Cancer* **53**, 727–732 (1986).
- Gendler, S. J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Wyman, A. & White, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6754–6758 (1980).
- Higgs, D. R., Goodbourn, S. E. Y., Wainscoat, J. S., Clegg, J. B. & Weatherall, D. J. *Nucleic Acids Res.* **9**, 4213–4224 (1981).
- Bell, G. I., Selby, M. J. & Rutter, W. J. *Nature* **295**, 31–35 (1982).
- Goodbourn, S. E. Y., Higgs, D. R., Clegg, J. B. & Weatherall, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4808–4812 (1983).
- Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67–73 (1985).
- Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **316**, 76–79 (1985).
- Stoker, N. G., Cheah, K. S. E., Griffin, J. R. & Solomon, E. *Nucleic Acids Res.* **13**, 4613–4622 (1985).
- Wong, Z., Wilson, V., Jeffreys, A. J. & Thein, S. L. *Nucleic Acids Res.* **14**, 4605–4616 (1986).
- Baird, M. *et al.* *Am. J. hum. Genet.* **39**, 489–501 (1986).
- Stahl, H.-D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 543–547 (1985).
- Manning, R. F. & Gage, L. P. *J. biol. Chem.* **255**, 9451–9457 (1980).
- Mustavitch, M. A. T. & Hogness, D. S. *Cell* **29**, 1041–1051 (1982).
- Azen, E. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **81**, 5561–5565 (1984).
- Edwards, Y. H. *et al.* *Ann. hum. Genet.* **48**, 119–127 (1984).

Localization of the receptor-binding protein adhesin at the tip of the bacterial pilus

Frederik Lindberg, Björn Lund, Lenore Johansson & Staffan Normark

Department of Microbiology, University of Umeå, S-901 87 Umeå, Sweden

Strains of the bacterium *Escherichia coli* that cause infections of the human urinary tract produce so-called Pap-pili, which are hair-like appendages consisting of about 10^3 helically arranged subunits of the protein PapA. These pili mediate binding to digalactoside-containing glycolipids present on the epithelial cells which line the urinary tract^{1,2}. Recently, it has been suggested that three proteins, PapE, PapF and PapG, are responsible for this binding^{3,4}. In the absence of PapA, non-piliated bacteria are formed which nonetheless exhibit binding, showing that the bulk of the pilus is not essential for binding^{3,5}. Although pili can form without PapF and PapG, such pili are unable to bind to the digalactoside³. The protein PapG mediates binding specificity in *trans*-complementation experiments, so this protein is the digalactoside-specific adhesin⁶. Using immuno-electron microscopy we have found that Pap-pili are heteropolymers composed of the major pilin, PapA, the minor pilins, PapE and PapF, and the adhesin, PapG. The last three proteins are located at the tip of the pilus.

Table 1 Titre of the different sub-sera in radioimmunoassays with various *pap* mutants

Strain	Genotype	Control	A	E	F	G	Adhesion
HB101/pPAP5	W	—	1:20,000	1:5,000	1:1,000	1:2,000	+
HB101/pPAP23	A ⁻	—	—	1:5,000	1:1,000	1:2,000	+
HB101/pPAP15	E ⁻	—	1:20,000	—	1:1,000	1:1,500	+
HB101/pPAP14	F ⁻	—	1:10,000	1:1,000	—	1:200	—
HB101/pPAP24	G ⁻	—	1:20,000	1:1,000	1:700	—	—
HB101/pPAP32	C ⁻	—	—	—	—	—	—
HB101/pPAP37	D ⁻	—	—	—	—	—	—

Data from refs 4 and 6. Bacteria were reacted with the various sub-sera followed by detection of bound radioactivity with ¹²⁵I-labelled protein-A. The wild-type genotype is denoted by 'W'. The titres given correspond to the dilution of antiserum giving half-maximal Protein-A binding. Titres of 1:20 or less are shown as '—'. Adhesion is measured as agglutination of human P₁-erythrocytes (carrying digalactoside) or digalactoside-linked latex beads³.

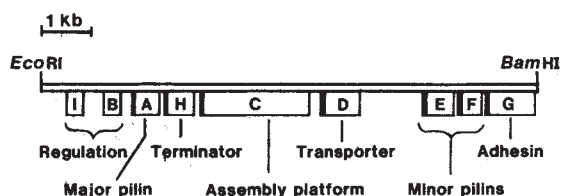


Fig. 1 Overview of the *pap* gene cluster on the 9.6-kilobase (kb) *EcoRI*-*Bam*HI fragment clones on plasmid pPAPS (ref. 3) with brief descriptions of the nine gene products are given. The product of *papA* is the major pilin¹⁹. PapH is pilin-like in sequence and seems to both terminate pilus growth and anchor the fully grown pilus to the cell surface⁹. The *papC* gene product is located in the outer membrane and forms the assembly platform for pilus growth²¹; PapD is a periplasmic protein which forms complexes with the pilins before assembly (unpublished data). PapE, PapF, and PapG are minor components which can be detected immunologically in the pilus. Both PapE and PapF are pilin-like in sequence⁴. PapE is also found in purified pili by SDS-polyacrylamide gel electrophoresis (PAGE)⁴. PapG is the adhesin as shown by the fact that the *papG* gene mediates binding specificity in *trans* complementation experiments⁶.

E. coli K12 strain HB101 (ref. 7) is non-piliated and does not bind to human erythrocytes or to other surfaces carrying digalactoside (galactose $\alpha(1 \rightarrow 4)$ galactose)-containing glycolipids^{3,5}. When the *pap* gene cluster from the uropathogenic *E. coli* strain J96 (ref. 8) was introduced into HB101 as the recombinant plasmid pPAP5 (Fig. 1), digalactoside-binding pili were produced³.

Specific antisera against the various Pap proteins were prepared by adsorption, with relevant *pap* mutants, of rabbit antiserum that was raised against purified whole Pap-pili from HB101/pPAP5 (refs 4 and 6). Serum obtained after repeated adsorption with a *papA*⁻ mutant was called A-serum. Analogous, E-, F-, and G-sera were also prepared. As a control, a negative serum was prepared by adsorption with the wild-type clone, HB101/pPAP5. Radioimmunoassays (RIA) reacting the sera with the surface of different *pap* mutants have been described previously^{4,6} and are summarized in Table 1. To localize the different Pap proteins on the cell surface, bacteria were incubated with antiserum, washed, incubated with gold-labelled protein A and then viewed under the electron microscope (Fig. 2).

In agreement with the fact that virtually all the pilus consists of PapA subunits, the A-serum reacted with the entire pilus specified by the wild-type clone, without preference for any portion⁴ (Fig. 2, A/W). In contrast, the E-serum bound almost exclusively to the tip of the pilus (Fig. 2, E/W). Controls using negative serum, or no serum at all, consistently yielded less than ten gold particles per cell, none of which displayed any preferential binding (data not shown). The same negative results were obtained when *PapC*⁻ or *PapD*⁻ mutants were used, as well as when any specific serum was tested with the corresponding mutant, for example, E-serum with a *PapE*⁻ mutant. Recently, the PapE protein was purified as a complex with the periplasmic transport protein PapD (unpublished data). Using a rabbit antiserum raised against this complex, the distribution of bound gold particles was identical to that found with E-serum (data not shown), verifying the specificity of the absorbed E-serum.

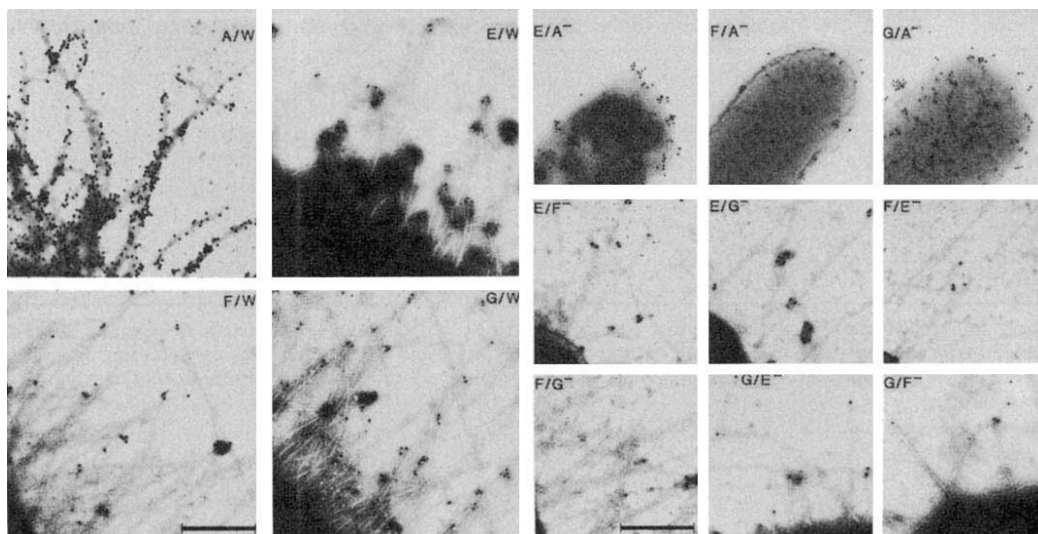
Both the PapF and PapG proteins were also found exclusively at the tip of the pilus (Fig. 2, F/W and G/W). In this case, fewer gold particles were bound than observed for PapE (Fig. 2, E/W), although almost all were attached to the tip. The sera were also used to examine the different mutant derivatives. In agreement with the RIA data (Table 1), PapE, PapF and PapG are expressed on the surface of a *PapA*⁻ mutant (Fig. 2, E/A⁻, F/A⁻ and G/A⁻). In the absence of any one of the proteins PapE, PapF, or PapG, the remaining two were still found at the tip of the pilus (Fig. 2).

From these results it is evident that *E. coli* Pap-pili are polymers of a major subunit, PapA, which accounts for the bulk of the pilus, and three minor subunits, PapE, PapF and PapG, which are located at the tip of the fibre. Another pilin-like protein, PapH, was recently shown to affect pilus length; it was postulated that PapH is a pilus-growth terminator that is incorporated as the last monomer at the pilus base⁹.

Figure 3 shows a model of Pap-pilus structure and biogenesis based on current knowledge. PapC is an outer-membrane protein which functions as a base for pilus polymerization²¹. The pilins PapA, PapE, PapF and probably PapH, are delivered to this platform as separate complexes with the periplasmic PapD protein (unpublished data). It is possible that PapG also forms a transient complex with PapD, although this has not yet been investigated. PapE has been detected in the purified pilus by silver staining of SDS-polyacrylamide gels. Under these conditions, PapF and PapG could not be identified with certainty, suggesting that PapE is present in greater amounts than the other two proteins⁴.

Unlike flagella¹⁰, we assume that Pap-pili grow from the base. Such an assembly mode has recently been demonstrated for the closely related type 1 pilus by pulse-labelling of growing pili with antibodies²². The first step of pilus assembly would be the association of PapG with the PapC platform, followed by the PapF subunit. Subsequently, a number of PapE subunits are polymerized into the pilus, possibly as one helical turn; $\sim 10^3$ PapA pilins are then incorporated. Finally, the addition of a

Fig. 2 Immuno-electron micrographs of different pap mutants reacted with sera specific for different Pap proteins. Each picture is labelled with the specificity of the serum as well as with the genotype of the bacterium. Thus, A/W is the A-serum reacted with the wild-type (HB101/pPAP5) and E/A⁻ is the E-serum reacted with the A⁻ mutant (HB101/pPAP23). The mutants and sera are as in Table 1. The left-hand portion of the figure shows wild-type cells reacted with the different sera; the right presents results of experiments with different mutants. Bars, 500 nm; all micrographs are of the same magnification.



Methods. Colloidal gold solutions were prepared by reduction of tetrachloroauric acid (HAuCl₄) with sodium citrate, as described by Frens²⁰. Briefly, 0.25 ml of 4% HAuCl₄ was added to 100 ml redistilled water. This solution was heated under reflux, 2.0 ml of 1% trisodium citrate was added and the solution boiled until dark red from colloidal gold. The pH was adjusted to 6.9 with 0.2 M K₂CO₃, and the solution filtered (0.45 µm) and stored at 4 °C. Protein-A (Pharmacia, 2.5 mg ml⁻¹) was added to 10 ml colloidal gold solution. After 3 min, 1 ml of 1% polyethylene glycol (PEG) was added and the mixture was centrifuged at 20,000g for 30 min at 4 °C. After another wash in PBS (50 mM potassium phosphate buffer at pH 7.4, 150 mM NaCl) supplemented with PEG at 0.2 mg ml⁻¹, the Protein A-gold complex was stored at 4 °C and diluted 10–20 times with PBS when required for use. Bacteria grown for 24–36 h at 37 °C on M9-Casa medium⁹ were resuspended in P-buffer (10 mM Tris-HCl at pH 7.2, 10 mM MgCl₂) and allowed to sediment onto a formvar-coated copper grid (400 mesh) for 10 min. The grid was then placed onto a drop of a 1:10 dilution of serum in RIA-buffer (PBS, 0.05% Brij 35, 0.05% NaN₃, 0.1% (BSA), 0.05% Triton X-100) for 3 min followed by washing (5 min) in a gently agitated microtitre well containing P-buffer. One drop of the solution containing colloidal gold-labelled protein-A was left on the grid for 3 min and the washing step repeated. Finally the preparations were negatively stained by a 10 s incubation in 1% sodium silicotungstate. All grids were examined in a Jeol 100-CX electron microscope operated at 80 kV.

PapH molecule is believed to terminate pilus growth, making the fibre complete⁹. The order of association of PapG and PapF in this model might be reversed.

Because PapE, PapF and PapG are at the pilus tip and never present in the body of the fibre, the order of incorporation of the different proteins must be strictly determined. Still, none of the three subunits seems to be required for the incorporation of the others, except that neither PapF nor PapG antigen can be detected in the absence of both PapA and PapE (Table 1). In the absence of PapF, piliation is drastically reduced³ suggesting an initiating, or rate-limiting, function for this protein in pilus assembly. In the absence of PapE, quantitatively and morphologically normal fibres are formed, but the association between the adhesin and the rest of the fibre is much weaker, as evidenced by the low binding-capacity of pili purified from a *PapE*⁻ mutant³. This is similar to the situation with PapH. In addition to its terminating function (shown by the longer pili on *PapH*⁻ mutant) the protein seems to be required for the secure anchoring of the fully grown pilus to the cell; pili detach much more easily from a *PapH*⁻ mutant than from the wild type⁹.

The finding that the PapG adhesin is located at one end of the pilus only, explains earlier findings that agglutination of human erythrocytes by pili, but not by whole bacteria, is greatly enhanced by Mg²⁺-mediated aggregation of pili³. Thus, to give multivalent agglutinating particles, pili have to be aggregated by solvent conditions or by their anchorage to bacteria.

At least 99% of the pilus is dispensable for attachment to erythrocytes and uro-epithelial cells⁵. Why then, is the adhesin associated with the pilus fibre? An attractive possibility is that the pilus is needed to place the adhesin outside the complete lipopolysaccharide O-chain normally found on uropathogenic *E. coli*. This suggestion¹¹ is based on the finding that such bacteria bind only if the *papA* gene is intact, although PapA itself is dispensable for binding of rough mutants lacking the O-side chain.

Biogenesis of the filamentous bacteriophages occurs in a similar manner to the model described here. Assembly starts

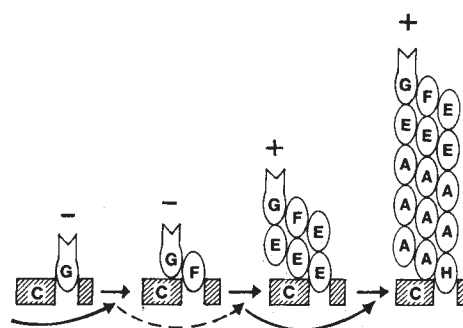


Fig. 3 Model for Pap-pilus structure and assembly. Four successive stages of assembly are shown from left to right. +, Indicates a Structure that can bind the digalactoside receptor; -, a structure unable to bind it (see the text for discussion).

with ten copies each of the C^{VII} and C^{IV} proteins followed by a large number of major capsid subunits¹². The other end is defined by five copies each of the A- and D-proteins. It is the A-protein that is the adhesin for future infection of other bacteria¹³. Furthermore, in the complicated T2 bacteriophage assembly systems, the adhesin, gp38, is also located at the tip—in this case, at the tip of the long tail fibres¹⁴.

In most other pilus systems essential accessory proteins have been found. At least in one case, the K88 system, the existence of a minor pilin has been postulated^{15,16}. In the Type 1 system, gene functions essential for adhesion but not required for piliation have been identified^{17,18}. Therefore, we believe that our conclusions are not limited to the Pap system, but rather that they represent a widely used solution to a common problem, namely the adhesion of pathogenic enterobacteria to the epithelial surfaces of their hosts.

We thank Drs Christopher Korch and Jan Tennent for critically reading this manuscript, and Eva Svenson for secretarial

assistance. We are also in debt to the Swedish Defence Research Institute, Umeå, Sweden, for the use of their electron microscope. This work was supported by grants from the Swedish Medical Research Council, the Swedish Natural Science Research Council, and the Swedish Board for Technological Development.

Received 24 February; accepted 1 May 1987.

1. Leffler, H. & Svanborg-Edén, C. in *Microbial Lectins and Agglutinins: Properties and Biological Activity* (ed. Mirelman, D.) 83-111 (Wiley, New York, 1986).
2. Normark, S. et al. in *Microbial Lectins and Agglutinins: Properties and Biological Activity* (ed. Mirelman, D.) 113-143 (Wiley, New York, 1986).
3. Lindberg, F. P., Lund, B. & Normark, S. *EMBO J.* **3**, 1167-1173 (1984).
4. Lindberg, F., Lund, B. & Normark, S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1891-1895 (1986).
5. Uhlin, B.-E., Norgren, M., Båga, M. & Normark, S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1800-1804 (1985).
6. Lund, B., Lindberg, F., Marklund, B.-I. & Normark, S. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
7. Boyer, H. W. & Roulland-Dussoix, D. *J. molec. Biol.* **41**, 459-472 (1969).
8. Hull, R. A., Gill, R. E., Hsu, P., Minshew, B. H. & Falkow, S. *Infect. Immunity* **33**, 933-938 (1981).
9. Båga, M., Norgren, M. & Normark, S. *Cell* **49**, 241-251 (1987).
10. Emerson, S. V., Tokuyasu, K. & Simon, M. I. *Science* **169**, 190-192 (1970).
11. Van Die, I., Zuidweg, E., Hoekstra, W. & Bergmans, H. *Microb. Pathogen.* **1**, 51-56 (1986).
12. Lopez, J. & Webster, R. E. *Virology* **127**, 177-193 (1983).
13. Grant, R. A., Lin, T.-C., Koningsberg, W. & Webster, R. E. *J. biol. Chem.* **256**, 539-546 (1981).
14. Riede, I., Drexler, K., Schwarz, H. & Henning, U. *J. molec. Biol.* **194**, 23-30 (1987).
15. Mooi, F. R., van Buuren, M., Koopman, G., Roosendaal, B. & de Graaf, F. K. *J. Bact.* **159**, 482-487 (1984).
16. Mooi, F. R., Roosendaal, B., Oudega, B. & de Graaf, F. K. in *Protein-Carbohydrate Interactions in Biological Systems*, (ed. Lark, D. L.) 19-26 (Academic, London, 1986).
17. Maurer, L. & Orndorff, P. E. *FEMS microbiol. Lett.* **30**, 59-66 (1985).
18. Mimion, F. C., Abraham, S. N., Beachey, E. H. & Goguen, J. D. *J. Bact.* **165**, 1033-1036 (1986).
19. Båga, M. et al. *J. Bact.* **157**, 330-333.
20. Frens, G. *Nature* **241**, 20-22 (1973).
21. Norgren, M., Båga, M., Tennent, J. M. & Normark, S. *Molec. Microbiol.* (in the press).
22. Lowe, M. A., Holt, S. C. & Eisenstein, B. I. *J. Bact.* **169**, 157-163 (1987).

A factor discriminating between the wild-type and a mutant polyomavirus enhancer

Imre Kovesi*, Masanobu Satake†, Kazuhiro Furukawa†, Ronald Reichel*, Yoshiaki Ito† & Joseph R. Nevins*

* Howard Hughes Medical Institute, Rockefeller University, New York, New York 10021, USA

† Institute for Virus Research, Kyoto University, Kyoto, Japan

Enhancers increase the frequency of transcription initiation from linked promoter elements¹, most probably as a result of the binding of specific proteins to the enhancer. The polyomavirus early region is expressed in differentiated mouse cells but not in undifferentiated embryonal carcinoma (EC) cells^{2,3}. This host range is a function of the enhancer because polyomavirus mutants selected for growth in EC cells have mutations in the enhancer⁴⁻⁶ and the host range is reproduced in transfection assays using the mutant enhancers⁷. To understand the basis for this alteration in enhancer function, we have assayed extracts of EC cells for proteins that can interact with this sequence. We have detected a protein, present in a variety of cells, that can bind to the F441 mutant sequence, but binds only very poorly to the wild-type sequence. We conclude that this sequence alteration has probably generated a binding site for a positive-acting factor that allows the enhancer to function.

Figure 1 shows a diagram of the polyomavirus enhancer, including the A and B elements as previously defined⁸, and the position of the alteration in the polyomavirus mutant F441. This mutation makes the virus able to grow in undifferentiated EC cells⁵. Also shown is the DNA sequence of oligonucleotides, representing the wild-type and F441 mutant sequence, synthesized to probe for specific proteins. The wild-type (WT) oligonucleotide and the mutant (F441) oligonucleotide were each cloned into the *Bam*HI site of the plasmid pUC13 and recovered by digestion with *Eco*RI and *Hind*III. The fragments were end-

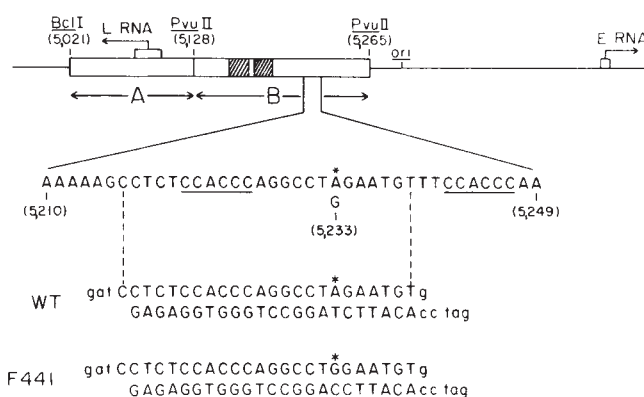


Fig. 1 Diagram of the polyomavirus enhancer depicting the previously defined A and B elements⁸ and the sites for interaction with nuclear proteins¹⁶⁻¹⁸, indicated by the hatched boxes. Below, the sequence of the region encompassing the point mutation in the F441 mutant virus (asterisk). Underline, sequences homologous to the BPV enhancer¹⁹ (the BPV homology). Oligonucleotides representing polyomavirus A2 wild-type sequence and the sequence of the F441 mutant are shown; these were cloned into the *Bam*HI site of pUC13.

labelled and used for binding measurements by the gel-shift assay^{9,10}. Using a nuclear extract prepared from undifferentiated F9 cells, we detected an interaction with the F441 mutant sequence, as evidence by the appearance of a major band of slower mobility (lane 1). In fact, two complexes (band A and band B) were obscured using the F441 probe. Both bands appeared to be due to specific interactions because they were eliminated by competition with an excess of the F441 sequence (lane 2). In sharp contrast, no such interactions could be detected when the WT probe was used for binding assays (lane 3), despite the fact that the two probes (F441 and WT) are identical with the exception of a single base pair. The specificity of binding is also demonstrated by the competition experiment, using the F441 DNA as a probe, shown in Fig. 2b. The homologous F441 sequence could effectively compete for binding to the F441 probe whereas the WT sequence was an inefficient competitor. Thus, we conclude that there is a factor or factors present in F9 cells that can specifically recognize the sequence in the F441 polyomavirus enhancer. The factor(s) appear to recognize the WT sequence but with an affinity much below that for the mutant sequence.

These results indicate that there is specific binding to the site of the F441 mutation. We have analysed the binding site by DNase footprinting. After a binding reaction and DNase digestion, complexes were resolved in an acrylamide gel, the A and B bands were eluted along with the free, unbound DNA, and each was analysed in a sequencing gel (Fig. 3a). There was indeed protection from DNase cleavage in the region encompassing the F441 mutation. Furthermore, the protection was identical for both the A and the B complex. The same was true when the complexes were analysed by methylation interference as shown in Fig. 3b. Methylation of a series of G and A residues within the region of DNase protection clearly inhibits binding of the factor in either the A complex or the B complex. Thus, either two forms of the same protein recognize the same site (for instance, the B complex could result from a modified form or a breakdown product of the protein that generates the A complex) or two distinct proteins recognize the same site.

We have also assayed for binding by direct footprint analysis without prior gel fractionation. Both the F441 probe and the WT probe were incubated with an F9 whole cell extract, treated with DNase and then analysed. The direct comparison of protection of the F441 probe and the WT probe reveals F441-specific-

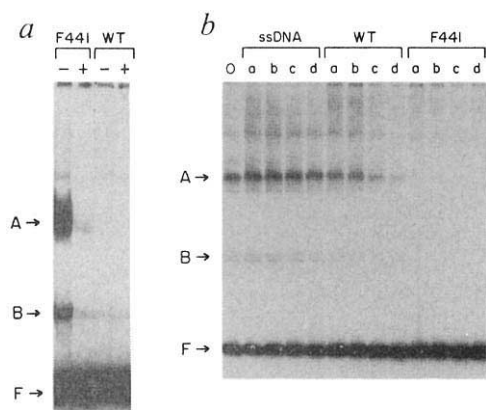


Fig. 2 Binding assay for factors in F9 cell extracts that interact with the wild-type and F441 enhancer sequence. *a*, A binding assay was performed as previously described²⁰ using a nuclear extract²¹ prepared from F9 cells with the exception that poly(dI)·poly(dC) was used as non-specific competitor rather than salmon sperm DNA. F441 and WT (above), origin of the labelled probe. Binding reactions were performed in the presence (+) or absence (-) of the unlabelled homologous fragment (30 ng). A, B, Specific complexes; F, free probe. *b*, Competition assay for specificity of binding. Binding assay as in *a*; in all cases, the labelled probe was the F441 fragment. Indicated at the top is the origin of the specific, unlabelled competitor DNA. Binding was carried out in the presence of 3 µg of poly(dI)·poly(dC) and in the absence (0) or presence of additional competitor which was either salmon sperm DNA (ssDNA), the homologous F441 sequence (F441), or the wild-type sequence (WT). Lanes a, b, c and d represent the addition of 30, 60, 90 and 150 ng of competitor, respectively.

Methods. Probes were prepared by digesting the WT and F441 oligonucleotide-containing pUC plasmids with *Eco*RI and *Hind*III to liberate the inserts. Probes were labelled at the *Eco*RI site with Klenow enzyme. In each case, 0.3 ng of ³²P-labelled probe was assayed with 10 µg of extract with 3 µg of poly(dI)·poly(dC). After 30 min at room temperature, binding reactions were analysed in a 6% acrylamide (0.1% bisacrylamide) gel followed by autoradiography.

protection in the same region protected in the gel-fractionated complex (Fig. 3c). In addition, there is protection in an adjacent region that includes the bovine papilloma virus (BPV) homology suggesting the binding of an additional factor when high concentrations of protein are used, as is necessary in this analysis. Because the difference in the protected region with F441 and WT protection includes the BPV region, it is possible that binding to the BPV site may require binding of the protein to the F441 mutant site. The difference in pattern of the wild-type DNA and the sample without extract is probably due in part to the difference in conditions of DNase digestion as a result of the extract although it is possible that there may be some protection of the WT sequence at this high protein concentration. Nevertheless, the clear result is the difference between the WT and F441 protection since these conditions are identical.

Finally, we have assayed for the F441-specific binding factors in extracts of a variety of other cell types including L929 cells, NIH 3T3 cells, mouse liver, Rev7 rat hepatoma, F9 cells, differentiated F9 cells and HeLa cells. There is binding activity in each cell type although the absolute level varies, as does the ratio of the A-complex to the B-complex. In addition, it is clear that there is little or no change in the level of the factor(s) as a function of differentiation of F9 cells. Therefore, we conclude that the factor or factors that recognize the F441 polyomavirus enhancer are not restricted to undifferentiated EC cells but rather can be found in a variety of cells.

The results presented here address the mechanism by which the single nucleotide change in the F441 mutant allows poly-

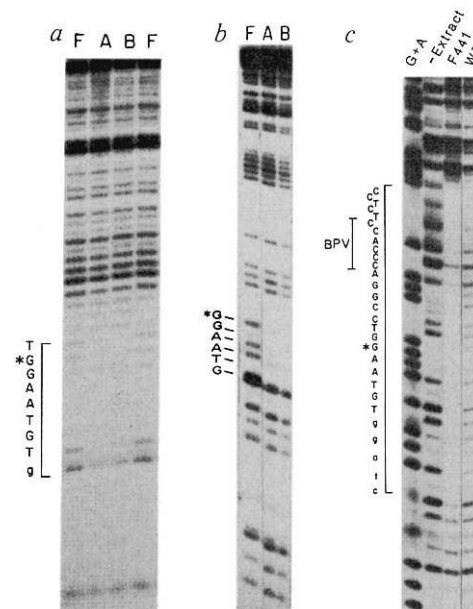


Fig. 3 Analysis of the binding of the F9 factor to the polyoma F441 sequence. *a*, A binding reaction was performed using the F441 probe and nuclear extract from F9 cells as in Fig. 2. After binding, the sample was treated with DNase I (ref. 22) and then analysed in a 6% acrylamide gel. The bands corresponding to complex A and B as well as free DNA were eluted, DNA was extracted, and then analysed in a 12% acrylamide-urea sequencing gel. The sequence of the DNA in the region of protection is indicated; *, F441 specific nucleotide. Sequence in lower case derives from vector. *a*, A complex; *b*, B complex; *f*, free DNA. *b*, Methylation interference assay. A binding reaction similar to that in panel *a* was carried out with partially methylated F441 probe. Complexes *a* and *b* were recovered from the gel and the DNA was analysed in a 12% acrylamide-urea sequencing gel after cleavage with alkali²³. *c*, Direct footprint analysis using crude whole-cell extract and no prior fractionation of complexes. End-labelled F441 probe and WT probe were incubated with whole-cell extract from F9 cells and then digested with DNase. Digested samples were analysed in a acrylamide-urea sequencing gel. The sequence of the DNA in the region of protection is shown. The BPV homologous sequence is indicated; *, F441 specific nucleotide. **Methods.** The procedures for DNase footprint analysis of DNA from gel complexes has been described previously²⁰. The procedure for methylation of probe DNA and analysis of binding interference were described²³. For the direct footprint analysis, 100 µg of whole-cell extract from F9 cells was incubated with ³²P-labelled F441 or WT probe in a 40 µl volume in the presence of 3 µg of poly(dI)·poly(dC). After 30 min incubation, 2 µg of DNase I was added and incubation continued for 2 min. The DNA was extracted and analysed in a 12% acrylamide-urea sequencing gel.

omavirus to function in undifferentiated EC cells. It might be imagined that such an alteration could either eliminate the binding of a negatively acting factor (repressor) or facilitate the binding of a positively acting factor by creating or improving a binding site. The data presented here suggest the latter. Apparently, the host range of the enhancer, and thus the virus, has been expanded as a result of the mutation by allowing an interaction with a factor that is found in a variety of cells. Furthermore, it may well be that the functional complex includes proteins bound to the BPV homology and that the binding of the factor to the F441 element is critical for the formation of the complex. The suggestion of a positive mechanism is consistent with the observation that the F441 base change is often duplicated in virus isolates^{5,6}, a duplication which also includes one or both of the BPV elements. If the mutation inactivated a repressor binding site then it would be difficult to imagine how

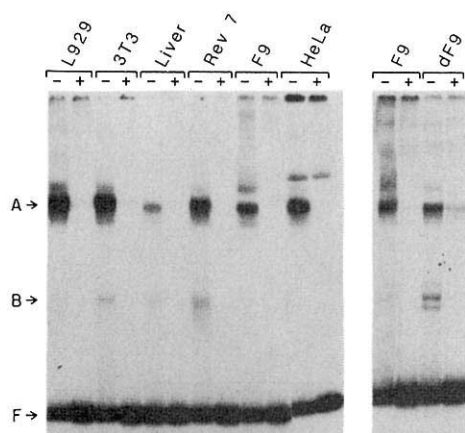


Fig. 4 Binding assays for F441-specific factor in extracts of various cell types. Binding assays using the F441 probe were carried out as in Fig. 2 using nuclear extract prepared from L929 cells, NIH 3T3 cells, mouse liver, Rev7 rat hepatoma cells²⁴, undifferentiated F9 cells, F9 cells induced to differentiate with cAMP and retinoic acid²⁵, and HeLa cells. For each extract, binding was performed without specific competitor (-) or with the homologous F441 mutant sequence (+).

there would be selective pressure to duplicate an inactivated site whereas the duplication of a positive binding site is a common occurrence (for instance, the early simian virus 40 (SV40) promoter which has five Sp1 binding sites¹¹).

The results presented here do not address the mechanism by which wild-type polyomavirus is restricted in F9 cells and functional in differentiated cells. Recent experiments suggest that at least a part of the restriction to SV40 and murine leukaemia virus (MLV) expression in EC cells is repression through the enhancer¹² which may be related to the negative action of products of the adenovirus *E1A* gene on enhancers¹³ and the presence of an *E1A*-like activity in F9 cells¹⁴. Moreover, a recent study has demonstrated that the F441 mutant enhancer is not repressed by the action of the *E1A* gene, whereas the wild-type enhancer is repressed, leading to the conclusion that the point mutation eliminated a repressor binding site¹⁵. Although our conclusions are at variance with those drawn by these authors, in fact the data are not contradictory. We would suggest that if such negative action is the primary reason for the inactivity of wild-type virus in F9 cells, then our results indicate that the negative action can be overcome by the acquisition of a binding site for a positive factor. Whether this means that the positive factor precludes the binding of a negative factor or possibly that both factors bind but the positive action is dominant is not settled by these experiments. In any case, the findings demonstrate that the interplay of multiple factors interacting with a genetic regulatory element can markedly affect the function of the linked gene. The results also underscore, in molecular terms, the dramatic consequences of a single point mutation on the regulatory properties of the cell.

We thank Dr Yutaka Teranishi, Laboratory of Biochemistry, Research Institute, Mitsubishi Kasei Kogyo Co. for the oligonucleotides. I.K. is a NSERC (Canada) postdoctoral fellow; R.R. is a Deutsche Forschungsgemeinschaft postdoctoral fellow. This work was supported by a grant from the NIH.

Received 25 March; accepted 29 April 1987.

1. Serfling, E., Jasin, M. & Schaffner, W. *Trends Genet.* **1**, 224-230 (1985).
2. Swartzendruber, D. E. & Lehman, J. M. *J. cell Physiol.* **85**, 179-188 (1975).
3. Boccardo, M. & Kelly, F. *Virology* **90**, 147-150 (1978).
4. Katinka, M., Yaniv, M., Vasseur, M. & Blangy, D. *Cell* **20**, 393-399 (1980).
5. Fujimura, F. K., Deininger, P. L., Friedmann, T. & Linney, E. *Cell* **23**, 809-814 (1981).
6. Sekikawa, K. & Levine, A. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1100-1104 (1981).
7. Linney, E. & Donerly, S. *Cell* **35**, 693-699 (1983).

8. Herbolmel, P., Bourachot, B. & Yaniv, M. *Cell* **39**, 653-662 (1984).
9. Garner, M. M. & Revzin, A. *Nucleic Acids Res.* **9**, 3047-3060 (1981).
10. Fried, M. & Crothers, D. M. *Nucleic Acids Res.* **9**, 6506-6525 (1981).
11. Gidoni, D., Dynan, W. S. & Tjian, R. *Nature* **312**, 409-413 (1984).
12. Gorman, C. M., Rigby, P. W. J. & Lane, D. P. *Cell* **42**, 519-526 (1985).
13. Borrelli, E., Hen, R. & Chambon, P. *Nature* **312**, 608-612 (1984).
14. Imperiale, M. J., Kao, H.-T., Feldman, L. T., Nevins, J. R. & Strickland, S. *Molec. cell Biol.* **4**, 867-874 (1984).
15. Hen, R., Borrelli, E., Fromental, C., Sassone-Corsi, P. & Chambon, P. *Nature* **321**, 249-251 (1986).
16. Piette, J., Kryszke, M.-H. & Yaniv, M. *EMBO J.* **4**, 2675-2685 (1985).
17. Fujimura, F. K. *Nucleic Acids Res.* **14**, 2845-2861 (1986).
18. Bohnlein, E. & Gruss, P. *Molec. cell Biol.* **6**, 1401-1411 (1986).
19. Weiher, H. & Botchan, M. R. *Nucleic Acids Res.* **12**, 2901-2916 (1984).
20. Kovsed, I., Reichel, R. & Nevins, J. R. *Cell* **45**, 219-228 (1986).
21. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475-1489 (1983).
22. Galas, D. & Schmitz, A. *Nucleic Acids Res.* **5**, 3157-3170 (1978).
23. Sen, R. & Baltimore, D. *Cell* **46**, 705-716 (1986).
24. Deschatreite, J., Moore, E. E., Dubois, M. & Weiss, M. C. *Cell* **19**, 1043-1051 (1980).
25. Strickland, S. & Mahdavi, V. *Cell* **15**, 393-403 (1978).

A single mode of DNA base-pair opening drives imino proton exchange

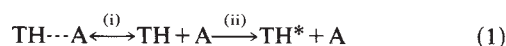
Maurice Guéron, Michel Kochoyan & Jean-Louis Leroy

Groupe de Biophysique, Ecole Polytechnique, 91128 Palaiseau, France

The opening of base pairs of double-stranded DNA is an important process, being a prerequisite for replication and transcription and possibly a factor in the recognition, flexibility and structure of DNA. The kinetics of base-pair opening have, however, been controversial. Base-pair opening can be studied by following the exchange of protons from imino groups with water, a process that seems only to occur from open base pairs. We have recently demonstrated catalysis by proton acceptors of imino proton exchange in nucleic acids. This has enabled us to determine the base-pair lifetimes, which are in the region of 10 ms at room temperature^{1,2}. In earlier reports it had been considered that proton exchange is limited by the rate of base-pair opening, which had led to estimates of base-pair lifetimes that were larger by one or two orders of magnitude³⁻¹⁷. There are also important discrepancies between recent and early estimates of the base-pair dissociation constant. Earlier estimates of base-pair lifetimes correspond in fact to the time required for proton exchange in the absence of added catalyst (AAC exchange). This could be a distinct mode of base-pair opening with a very long open lifetime, different from the mode revealed by the effect of catalyst. The evidence reported here suggests on the contrary that there is only a single mode of base-pair opening and that proton exchange in the absence of added catalyst is in fact catalysed by a proton acceptor intrinsic to the nucleic acid, most probably the other base of the open pair.

Base-pair lifetime can be determined by following proton exchange. Proton exchange is catalysed by proton acceptors, such as H₂O and OH⁻, and the base component of buffers³, such as ammonia, imidazole, Tris and so on. For instance, exchange of the imino proton of the mononucleoside thymidine at pH 6 in water is catalysed by OH⁻ and occurs in 8 ms at 25 °C. Ammonia base at 1 μM catalyses exchange in 4 ms (observed at pH 5, manuscript in preparation).

Imino proton exchange from a base pair (for example, an A·T pair) occurs in two steps, namely pair opening (i) followed by proton exchange from the open state (ii) (A and T are conventional nucleotide symbols, H represents the proton concerned):



If the catalyst concentration is large enough, exchange occurs for each opening event, and the exchange time is equal to the base-pair lifetime. If the concentration is low, many openings are required: assuming that exchange from the open state occurs

as from the isolated nucleoside, the exchange time is equal to that of the monomer in the same conditions of catalysis, divided by the base-pair dissociation constant^{1,5}.

The general formula for the exchange time τ_{ex} is:

$$\tau_{ex} = \tau_0 + \tau_i / (K\alpha) \quad (2)$$

where τ_0 and K are the base-pair lifetime and dissociation constant respectively, and τ_i is the exchange time for the mononucleoside, which is proportional to the inverse of the catalyst concentration. The coefficient α ($\alpha \leq 1$) takes into account hindered access of the catalyst to the imino proton of the open base pair, as compared to the mononucleoside. The comparison of the efficiency of different exchange catalysts suggests that α is not much smaller than unity (ref. 2, and manuscript in preparation). Thus, imino proton exchange measurements—most conveniently carried out by proton magnetic resonance (PMR) spectroscopy—can provide both the lifetime and, less directly, the dissociation constant of a base pair.

In the conditions customary in nucleic-acid studies, proton exchange is limited by the catalytic step rather than by base-pair opening^{1,2}. This is easily demonstrated by increasing the concentration of catalyst, as shown in Fig. 1 for the A·T base pair of the self-complementary oligodeoxynucleotide, 5'-CGCGATCGCG. The plot of the imino proton exchange time against the inverse of the concentration of NH_3 catalyst is a straight line, as predicted by equation (2). Extrapolation to infinite catalyst concentration yields a base-pair lifetime of 2 ± 0.7 ms at 25 °C. Turning to a lower ammonia concentration, for example 20 mM, the exchange time is 19 ms, compared to that of the monomer in $1 \mu\text{M}$ NH_3 , namely 4 ms. Hence the dissociation constant of the base pair is (assuming $\alpha = 1$):

$$(4/19) \times 10^{-6} / (20 \times 10^{-3}) \approx 10^{-5}$$

Similar values are found for A·T base pairs in different duplexes. The lifetimes of G·C pairs are typically three times longer, and their dissociation constants are 10 times smaller. Neighbouring base pairs have different lifetimes, hence base pairs open one at a time², as predicted earlier^{27,28}. In poly(rA)·poly(rU) the dissociation constant is¹ in the region of 10^{-3} . Catalysed exchange has also been observed in transfer RNA^{18,19} and in DNA (P. Bendel, personal communication).

Surprisingly, the exchange time remains finite even in the absence of added catalyst (the AAC exchange process). Exchange of this type is unaffected by catalysts added in low concentrations and is also independent of pH when the pH is not too high (or too low, in the case of guanosine; see Fig. 3). Hence there is a threshold for the effect of added catalysts, which explains why the effect had been missed previously. Such exchange has been observed repeatedly by PMR over the last decade, in DNA and RNA duplexes^{7,8,10-15}, as well as in RNA^{20,21}. Two explanations may be considered.

The first refers to base-pair kinetics. Concurrently with the kinetic process characterized above (process 1), which we consider to be incontrovertible, a second process is postulated, in which opening would occur much less often, namely every 260 ms at 25 °C for the A·T pair of Fig. 1, and the open state would be so long-lived that the imino proton would exchange at each opening event. This is the original interpretation of AAC exchange. For the A·T pair of Fig. 1, the catalyst at pH 6 could be OH^- , and an open-state lifetime larger than 8 ms would be required, corresponding to a dissociation constant larger than 3×10^{-2} . At pH 5, the catalyst would be H_2O , with corresponding values of 20 ms and 8×10^{-2} . Together, the two kinetic processes could account for all the exchange data of Fig. 1. However the large dissociation constant required by the second process is in conflict with values based on the kinetics of denaturation by formaldehyde²² or on hydrodynamic properties²³ of DNA. The same problem arises with the kinetics of mercury binding to poly(rA)·poly(rU) (ref. 24).

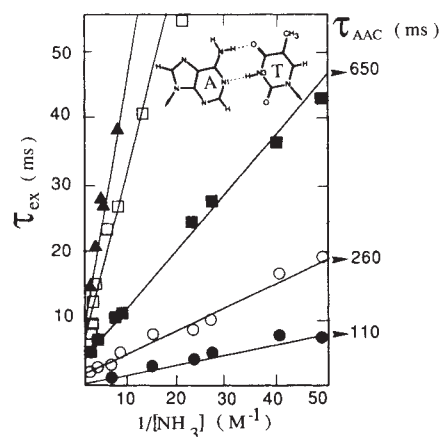


Fig. 1 The exchange time τ_{ex} of the imino proton of thymidine of the self-complementary oligodeoxynucleotide 5'-CGCGATCGCG in the duplex form plotted against the inverse of the concentration of the NH_3 used as a catalyst; 0 °C ($\blacktriangle$), 5 °C ($\square$), 15 °C ($\blacksquare$), 25 °C ($\circ$) and 35 °C ($\bullet$). The exchange time varies with catalyst concentration and extrapolates at each temperature to a finite value, the base-pair lifetime, for infinite catalyst concentration. For moderate concentrations, the exchange is catalyst-limited, and the base-pair dissociation constant is obtained as the ratio of the exchange time of isolated thymidine to that of the duplex. At very low NH_3 concentrations, the exchange time levels off at the value given on the right-hand side, τ_{AAC} for each temperature. Such exchange in the absence of added catalyst (AAC) is ascribed to catalysis by adenosine in the open state (see text). **Methods.** Proton exchange with water through ammonia was measured by ammonia-induced variations of line width, of longitudinal relaxation or of the rate of magnetization transfer from water. Catalyst-independent exchange was measured by magnetization transfer. The solution contained: duplex, 1 mM; NaCl, 100 mM, and ammonium buffer in the appropriate amount to obtain the stated ammonia concentration. The buffer concentration was measured by the intensity of its proton spectrum in a sample aliquot at pH 2. The sample pH was 8.8. The proton frequency was 276 MHz.

In the other explanation, the model of base-pair kinetics is strictly the one derived from the catalyst-dependent exchange measurements (process 1). The AAC exchange is ascribed to an intrinsic catalyst. Because there is no evidence for exchange from closed base pairs, exchange in the absence of catalyst must occur from the open state of process 1. Remarkably, this proposition lends itself to a decisive experimental test. If exchange without catalyst operates on the same open state as added catalysts such as ammonia, the rates of exchange by the AAC mechanism and by ammonia catalysis at a given concentration should be in a constant ratio.

This prediction is borne out by the data. In Fig. 1, the AAC exchange times at 35, 25 and 15 °C may be compared to the exchange times in 20 mM ammonia (50 M^{-1}). The ratios are respectively 15.7, 15.3 and 16.7. In other words, for this A·T pair, the AAC exchange mechanism is as efficient as exchange catalysis by 1.25 mM ammonia, and this relationship is maintained in a range of temperatures where the efficiency of either changes by more than a factor of five. (Other catalysts can be substituted for ammonia, with similar results).

The comparison of the AAC mechanism and of catalysed exchange has been carried out for other base pairs. The results are shown in Fig. 2, which includes all the base pairs for which we have been able to do the required exchange time measurements (at least four per base pair). For a constant ratio, all the points relative to one base pair would lie on a straight segment, of slope unity. Such a segment has been traced for each base pair, and one can examine how well the points are aligned on it.

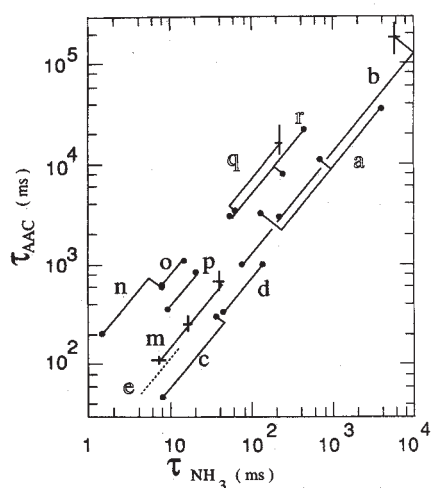


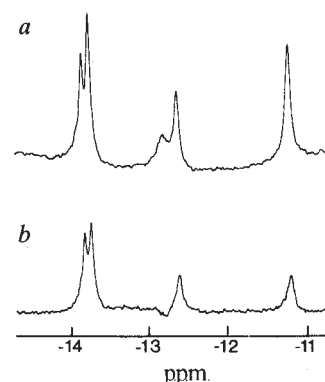
Fig. 3 Imino proton exchange of the terminal T (-12.7 p.p.m.) of d-AATTGCAATT, compared to the mononucleoside thymidine (-11.2 p.p.m.) in the same sample. The duplex and monomer concentrations are 0.25 mM and 1 mM respectively; pH 5, $T = 15^\circ C$. *a*, Normal spectrum. Exchange broadening is larger for terminal T than for the monomer. *b*, Spectrum at 12 ms following inversion of water proton magnetization. The peak of terminal T is already negative, whereas that of the monomer has decreased by 70% . From these data and others taken at different times, exchange times of 9 and 47 ms respectively were obtained. This demonstrates intrinsic catalysis of imino proton exchange. (The exchange of the G imino proton at -12.5 p.p.m. will be discussed elsewhere. It is acid-catalysed, owing to protonation on N7, an observation which supports our general interpretation.) The chemical shift reference is DSS (2,2-dimethylsilapentane-5-sulphonate).

The agreement is excellent for most of the base pairs which are beyond the third from the end of the duplex, namely those labelled b, d, m, o, p, q, r. Note that the upper point for base pair b corresponds to an AAC exchange time of three minutes ($T = 0^\circ C$), which was measured by hydrogen-deuterium exchange PMR.

Among the base pairs in the third position, c behaves as predicted, but the agreement is not as good for the two others, a and n. This may be due to end effects and would be explained by a temperature dependence of the structure of the open state. More generally, the prediction of a constant ratio assumes that the temperature dependence of exchange times is due mainly to the variation of the base-pair dissociation constant, and that the variation of other factors affecting exchange, such as the pK s of the proton donor and acceptors, or the diffusion of ammonia, has modest effects on the ratio. This is indeed what is suggested by preliminary measurements and computations. All things considered, the data of Fig. 2 strongly suggest that catalysis in the open state of process 1 is the explanation for AAC exchange.

There is direct evidence for intrinsic catalysis in the case of the terminal thymidine of the self-complementary DNA duplex AATTGCAATT. The exchange time was measured in a sample which also contained the nucleoside thymidine. No catalyst was added. Because the availability of an imino proton for exchange cannot be larger in a duplex than in the mononucleoside, the

Fig. 2 Comparison of exchange in the absence of added catalyst and exchange due to added catalyst. The abscissa is the exchange time in 20 mM ammonia, minus the base-pair lifetime. The ordinate is τ_{AAC} , the exchange time in the absence of added catalyst; subtraction of the base-pair lifetime is a minor correction. The measurements with and without catalyst were carried out as in Fig. 1. The dotted line corresponds to data¹ for poly(rI)·poly(rC), catalysed by Tris base. For each base pair, a straight line of slope unity is drawn through the points corresponding to measurements at two or three temperatures. The fit shows that the efficiencies of the two exchange mechanisms are proportional. The inference is that exchange without added catalyst takes place in one and the same open state of the base pair as exchange catalysed by NH_3 . The ratio of efficiencies differs according to the base pair, as shown by the non-alignment of the different segments on the log-log plot. The symbols represent particular base pairs as follows: a, b, m, the third fourth and fifth base pairs in $5'$ -CGCGATCGCG; c, the third base pair in CG5mCGCG(B-form); n, o, p, the third, fourth and fifth base pairs in CCTTCGAAAGG; q, r, d, the fourth, fifth and sixth base pairs in GGAAAGCTTTCC; e represents base pairs of poly(rI)·poly(rC). The measurements were carried out at the following temperatures: for a, m and r, at $35^\circ C$, $25^\circ C$ and $15^\circ C$; for c, d and e at $35^\circ C$ and $25^\circ C$; for n, o, p and q at $25^\circ C$ and $15^\circ C$ and for b at $35^\circ C$, $25^\circ C$ and $0^\circ C$. Letters in outline represent values scaled up by a factor of ten in both directions to improve readability (a, e, q and r). Error bars are given in a few representative cases.



exchange time should be longer for the imino protons of the former, including the imino proton of the terminal base pair. On the contrary the exchange time for the terminal T was 9 ms, five times shorter than for the monomer (47 ms), demonstrating intrinsic catalysis (Fig. 3).

In the absence of added catalyst, the catalyst must be a proton acceptor of the oligomer itself. The best candidate is the imido N1 of adenosine (or the cytidine N3 for a G·C pair). Its pK is 3.7 so that the efficiency for proton abstraction from thymidine ($pK = 9.8$) is low, close to 10^{-6} . Let us consider the A·T pair of Fig. 1 with its dissociation constant of 10^{-5} . During a period equal to the AAC exchange time of 0.26 s, the cumulative time spent in the open state is 0.26×10^{-5} s and there must be 10^6 collisions during that time. Hence the required collision rate in the open state is:

$$10^6 / (0.26 \times 10^{-5}) \approx 4 \times 10^{11} \text{ s}^{-1}$$

This is a very high rate. Furthermore, proton abstraction does not always lead to exchange with water. Indeed, direct collisions between the AH^+ and T^- species formed by proton abstraction should result in efficient back-transfer of the proton to thymidine. However, indirect back-transfer, through a water molecule, would lead to net exchange with a water proton. Starting from the pioneering work of Grunwald *et al.* on methyl ammonium²⁵, indirect transfer has been extensively studied and shown to be efficient (the so-called Grotthuss chain process)²⁶.

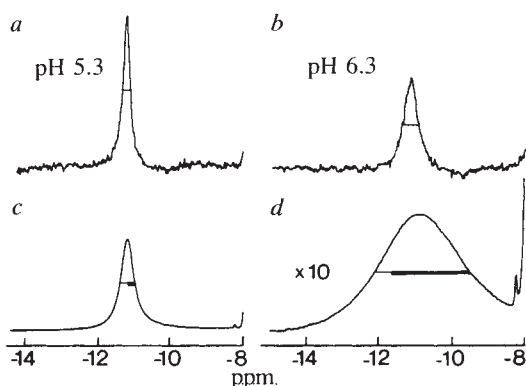


Fig. 4 Imino proton transfer from uridine (represented here by UH) to water, measured by line-broadening of the imino proton spectrum. *a, b*, The uridine concentration is 10 mM. Proton transfer is catalysed by H_2O at pH 5.3 and OH^- at pH 6.3. The concentration of U^- ion is too small to affect proton exchange. *c, d*, The uridine concentration is 1.16 M. The concentration-dependent increments of the linewidth are 50 and 558 Hz. Their ratio is close to the ratio (1:10) of the pH-dependent U^- concentrations, showing that the incremental width, represented by a heavy line, is due to indirect proton transfer to water through a $\text{UH}\cdots\text{OHH}\cdots\text{U}^-$ complex. The spectra are scaled according to the uridine concentration. The temperature was 25 °C.

Lacking an experimental model of indirect back-transfer for the AH^+/T^- couple, we illustrate the transfer to water by the homologous couple UH/U^- (Fig. 4).

For a large collision rate, the two bases should remain close to one another in the open state. Together with the requirement for indirect back-transfer, this suggests that an open-state structure in which hydrogen-bonded water bridges the N_1 of adenosine to the $-\text{N}_3\text{H}$ of thymidine should be considered. Preliminary evaluations indicate that such a structure could account for the observed AAC exchange times.

It was mentioned above that the ratio between the rates of exchange by the AAC mechanism and by ammonia catalysis will be constant only if the chemistry of exchange itself is not strongly temperature-dependent. This can be explored theoretically, by considering the temperature dependence of pK and of diffusion rates, or experimentally, by measuring exchange catalysis of mononucleosides at different temperatures, as will be reported in detail elsewhere. We point out here that the efficiency of ammonia as a catalyst of nucleosides is nearly independent of temperature, due to the compensating variations with temperature of the difference between its pK and that of nucleosides and of its diffusion constant. As for the pK difference between adenosine and thymidine, its variation should change the rate of catalysis by a factor of ~ 1.2 over 10 deg C, and the effect is even smaller for a G·C pair. Such variations are too small to affect the interpretation of the data of Fig. 2.

The identification of the other base of the pair as the intrinsic catalyst is, in our opinion, plausible but not yet established. However this does not alter the strong evidence presented above that ammonia-catalysed exchange and AAC exchange proceed from the same open state.

In conclusion, a single kinetic process is responsible for all features of imino proton exchange in the deoxyribonucleic acids studied here. Exchange occurs from the open state. It is catalysed both by added catalysts such as ammonia and by an intrinsic catalyst, probably the other base of the pair. In the latter case, proton transfer to water would occur during proton back-transfer through a bridging water molecule. The long catalyst-independent exchange time observed at low catalyst concentrations is unrelated to base-pair lifetime. We propose the present model

for the interpretation of imino proton exchange in A, B and Z duplexes and in transfer RNA.

This process may be the principal mode of base-pair disruption, or even the only one, at temperatures well below melting. Its properties are presently under investigation and some characteristic results have been presented earlier^{1,2}. At 25 °C, lifetimes are typically in the region of 10 ms. Dissociation constants are in the region of 10^{-5} for oligo-deoxyribonucleic acids. Internal base pairs open one at a time and their lifetime shows some sensitivity to the nature of the base pair and to the DNA sequence.

We thank Drs T. Huynh-Dinh and J. Igolen (Unité de Chimie Organique, Institut Pasteur), Dr G. Lancelot (Centre de Biophysique Moléculaire, Orléans) and Professor D. J. Patel (Columbia University, New York) for the oligodeoxynucleotides used in this work; Professor J. Delpuech (Université de Nancy) for discussion; and A. Moraillon for help with the experiments. This work was supported in part by a grant from 'Action Concertée: Interface Physique-Biologie', Fonds de la Recherche et de la Technologie.

Received 29 December 1986; accepted 28 April 1987.

1. Leroy, J. L., Broseta, D. & Guéron, M. *J. molec. Biol.* **184**, 165-178 (1985).
2. Guéron, M., Kochoyan, M. & Leroy, J. L. in *Proc. First EBSA Workshop* Saltsjöbaden, July 1986 (eds Ehrenberg, A., Rigler, R., Gräslund, A. & Nilsson, L.) 196-200 (Springer, Berlin, 1987).
3. Teitelbaum, H. & Englander, S. W. *J. molec. Biol.* **92**, 55-78 (1975).
4. Teitelbaum, H. & Englander, S. W. *J. molec. Biol.* **92**, 79-92 (1975).
5. Mandal, C., Kallenbach, N. R. & Englander, S. W. *J. molec. Biol.* **135**, 391-411 (1979).
6. Nakanishi, N., Mitane, Y. & Tsuboi, M. *Biochim. biophys. Acta* **798**, 46-52 (1984).
7. Mirau, P. A. & Kearns, D. R. *J. molec. Biol.* **77**, 207-227 (1984).
8. Mirau, P. A. & Kearns, D. R. *Biopolymers* **24**, 711-724 (1985).
9. Hartmann, B., Leng, M. & Ramstein, J. *Biochemistry* **25**, 3073-3077 (1986).
10. Patel, D. J. & Hilbers, C. W. *Biochemistry* **14**, 2651-2656 (1975).
11. Early, T. A., Kearns, D. R., Hillen, W. & Wells, R. D. *Biochemistry* **20**, 3756-3764 (1981).
12. Pardi, A. & Tinoco, I. Jr *Biochemistry* **24**, 4686-4693 (1982).
13. Chou, S.-H., Wemmer, D. E., Hare, D. R. & Reid, B. R. *Biochemistry* **23**, 2257-2262 (1984).
14. Patel, D. J., Kozłowski, S. A., Weiss, M. & Bhatt, R. *Biochemistry* **24**, 936-944 (1985).
15. Lefèvre, J. F., Lane, A. N. & Jardetzky, O. *J. molec. Biol.* **185**, 689-699 (1985).
16. Englander, S. W. & Kallenbach, N. R. *Q. Rev. Biophys.* **16**, 521-655 (1984).
17. Cantor, C. R. & Schimmel, P. R. *Biophysical Chemistry Part III* Ch. 22 (Freeman, San Francisco, 1980).
18. Figueroa, N. et al. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4330-4333 (1983).
19. Leroy, J. L., Bolo, N., Figueroa, N., Plateau, P. & Guéron, M. *J. biomolec. struct. Dynam.* **2**, 915-939 (1985).
20. Crothers, D. M., Cole, P. E., Hilbers, C. W. & Shulman, R. G. *J. molec. Biol.* **87**, 63-88 (1974).
21. Johnston, P. D. & Redfield, A. G. *Nucleic Acids Res.* **5**, 3913-3927 (1978).
22. Frank-Kamenetskii, M. D. in *Structure and Motion: Membranes, Nucleic Acids and Proteins* (eds Clementi, E., Corongin, G., Sarma, M. H. & Sarma, H.) 417-432 (Adenine Press, Guilderland, New York, 1985).
23. Wilcoxon, J. & Schurr, J. M. *Biopolymers* **22**, 2273-2321 (1983).
24. Williams, M. N. & Crothers, D. M. *Biochemistry* **14**, 1944-1951 (1975).
25. Grunwald, E., Lowenstein, A. & Meiboom, S. *J. Chem. Phys.* **27**, 630-640 (1957).
26. Bianchini, B. et al. in *Protons and Ions Involved in Fast Dynamic Phenomena* (ed. Lazlo, P.) 265-282 (Elsevier, Amsterdam, 1978).
27. Gralla, J. & Crothers, D. M. *J. molec. Biol.* **78**, 301-319 (1973).
28. Lukashin, A. V., Vologodskii, A. V., Frank-Kamenetskii, M. D. & Lyubchenko, Y. L. *J. molec. Biol.* **108**, 665-682.

Corrigendum

The optical polarization of the Sun, measured at a sensitivity of parts in ten million

J. C. Kemp, G. D. Henson, C. T. Steiner, I. S. Beardsley & E. R. Powell

Nature **326**, 270-273 (1987).

In this paper I. S. Beardsley was omitted from the published list of authors.

Erratum

Quasicrystal morphologies

Nature **326**, 16 April (1987).

THE cover photographs of this issue should be credited to J. M. Lang.

Summer science product debut

Molecular modelling software, an animal activity monitor, laboratory robotics and a database of translated Japanese journals round out this week's selections.

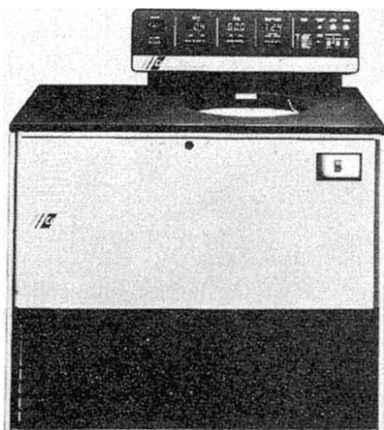
THE non-intrusive **animal activity monitor** from HVS Image Analysing Ltd uses an overhead TV camera to help researchers keep track of the movements of laboratory animals (Reader Service No. 100). The monitor can be used with a single animal, fish or insect of any colour. No special markers or lights are required, because the monitor works by picking up on the contrast of the animal against the



The animal activity monitor from HVS Image Analysing Ltd helps researchers keep track of their subjects.

background. The image gathered by the UV camera passes to the scanning unit, which HVS says can extract and locate the coordinates of an animal at rates of up to 50 times per second. The scanning unit interfaces with BBC, Apple or IBM PCs for data analysis and storage, and HVS supplies basic demonstration software that can be tailored to suit specific requirements. HVS sells its monitoring system, including scanning unit, interface and software, for \$7,000 (US), and says that it will have a monitor that can track two animals at a time on the market soon.

A basic, no-frills **ultracentrifuge** is new to Du Pont's line of Sorvall OTD centri-

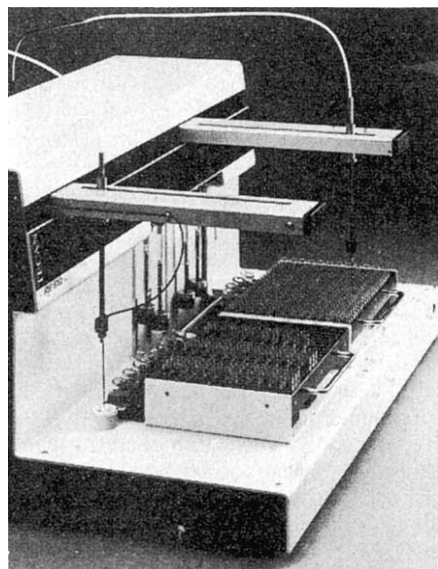


Du Pont has a new Sorvall ultracentrifuge, guaranteed for 30,000 million revolutions.

fuges (Reader Service No. 101). The Combi spins at one speed up to an 80,000 r.p.m. maximum, and can be set for differing temperatures and times, making it a workhorse for any laboratory. The Combi's wipe-out chamber eliminates the need for an intermediate dry cycle that can use up spin time unnecessarily. Du Pont sells the Combi for £19,500 (UK), and backs it up with a warranty for the drive system for up to 30,000 million revolutions.

The immunological response

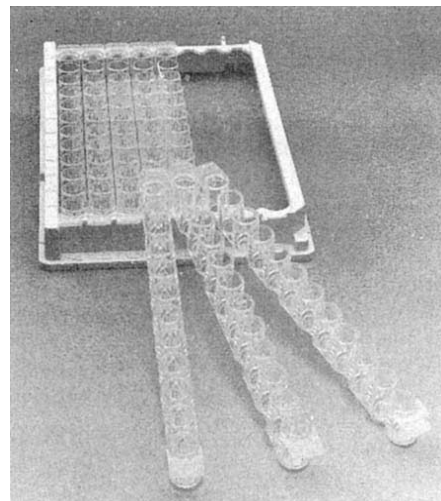
Tecan Ltd has a new **robotic sample processor** with an interesting structural feature — it has two arms (Reader Service No. 102). The RSP 5052's arms can work in concert or perform independent tasks, and each can be equipped with a single, dual or multi-tip pipette. The processor's



Two arms double the processing speed of Tecan Ltd's robotic sample processor.

capabilities can double lab throughput; with a single-tip pipette, one arm can carry out sample transfer, while the other dispenses reagents through a multi-tip. Special features include three tip wash stations, a flexible rack system, peristaltic and vacuum pumps, temperature control, and the ability to incorporate several diluter dispensers. The software is also flexible, allowing the user to add new control routines, and an integrated database can be used to create customized data output.

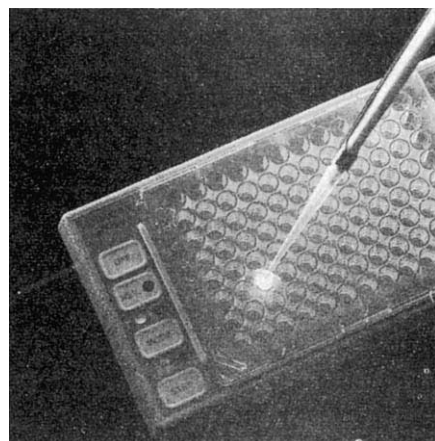
Nunc has designed a new **immunostrip** with wells that have sloped sides, so researchers no longer have to worry about not getting the nooks and crannies of



The individual strips snap out of Nunc's round-bottomed microplates.

microplates clean (Reader Service No. 103). Nunc's strips have twelve wells, and fit into self-standing, standard-sized racks that can be run on most sample processors and readers. The strips are composed of polystyrene, and Nunc says it has recorded successful results using the wells on both monochromatic and bichromatic readers. The modules, consisting of eight strips and a rack, are sold in cases of 60 plates.

With the TrakWell marker from FMC BioProducts, researchers may never lose their place on 96-well microplates again (Reader Service No. 104). TrakWell consists of an LED light unit that fits most standard microplates. It **illuminates each well** in which sample is being added or



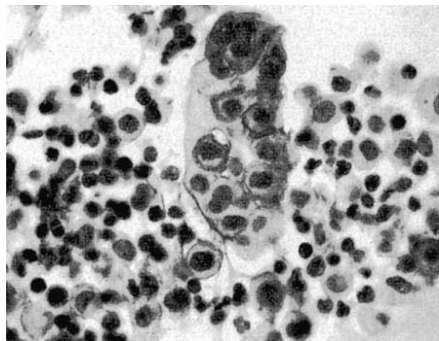
The TrakWell marker really shines a light on the subject.

removed, enabling workers to keep track of where they are. A foot-switch provides hands-free operation to move the light to

the next well. The TrakWell carries a \$325 (US) price tag.

IQ(Bio) Ltd has just launched a new line of **immunoenzymetric assays** that use the much-vaunted "sandwich assay" technique (Reader Service No. 105). The tests use two antibodies, one attached to a solid phase, and a second monoclonal antibody covalently linked to alkaline phosphatase, that trap the substance under study between them. After unbound antibody is washed off, the antibody bearing alkaline phosphatase dephosphorylates NADP to NAD, and a subsequent redox reaction reduces a colourless precursor to a formazan dye. IQ(Bio) reports a sensitivity for the assays of 0.01 mIU l^{-1} , and a working range (CV less than 10 per cent) of 0.1 to 25 mIU l^{-1} . The first product in this line is a kit for TSH, costing £70 (UK) for 96 determinations, but IQ(Bio) says assays for thyroxine, LH, FSH, prolactin and progesterone will be available soon.

Biomedical Technologies Inc. has a monoclonal antibody to detect tumour associated antigen (TAG-72) selectively expressed in adenocarcinomas of the breast, lung, ovary and colon (Reader Service No. 106). MAb B72.3 is an



Biomedical Technologies' MAb B72.3 reacting with a pleural effusion ductal carcinoma of the breast.

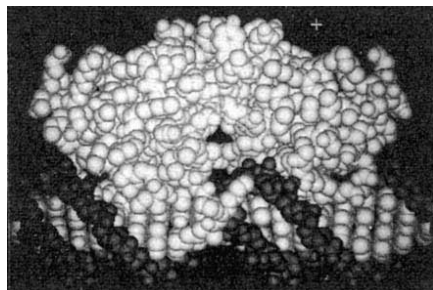
immunocytochemical adjunct for the **diagnosis of carcinoma** in cell blocks and cytocentrifuge preparations of formalin-fixed, paraffin-embedded human serous effusions. The antibody has also been shown to be useful for detecting "non-small cell" carcinomas in tissue gathered from fine needle aspiration biopsies of lung and other body sites, and adenocarcinomas of the breast and colon. Biomedical Technologies supplies MAb B72.3 in 100-test samples, complete with an immunohistochemical protocol, for \$200 (US).

From Immunodiagnostics Ltd comes a **magnetic immunoassay diagnostic kit** for prolactin (Reader Service No. 107). The Optimized MagNetic ImmunoAssay (OMNIA) system uses affinity-purified polyclonal antibody for prolactin bound to magnetic particles to make a one-step

assay that can be performed in one hour. To use the kit, $100 \mu\text{l}$ of sample is incubated with $200 \mu\text{l}$ of pre-mixed reagent containing radiolabelled anti-prolactin and magnetic anti-prolactin. The tubes are then placed on the OMNIA separator, available separately from Immunodiagnostics, for two separation steps. The typical resulting precision profile shows a variation of less than 10 per cent CV from 18 to $6,000 \text{ mIU l}^{-1}$, and less than 5 per cent from 45 to $6,000 \text{ mIU l}^{-1}$ according to the company. This wide range allows patients with hypoprolactinaemia to be better monitored throughout their therapy, and eliminates the need to dilute samples. The kits come in 100, 200 and 400 tube sizes, costing £75, £130 and £240 (UK), respectively.

Modelling and more

New England BioGraphics has a **macro-molecular graphics system**, priced to put



New England BioGraphics' space-filling model of the lambda repressor and operator (B-DNA).

molecular modelling within the reach of most laboratories (Reader Service No. 108). Version 2.0 of Promodeler I has just been released, and it allows proteins and nucleic acids to be displayed on an IBM AT or compatible equipped with the IBM Professional Graphics adapter and display, or equivalent. Promodeler I can manipulate up to 5,000 atoms simultaneously, including stereo views, in space filling, bond, and backbone representations. In bond or backbone mode, a molecule can be rotated, translated or scaled in "semi-real time", at a rate of 1–3 sec per position. Part or all of the molecule may be coloured according to colours selected from pop-up menus by a mouse. Subsets within a molecule may be coloured, rotated or translated independently, and portions of the molecule can be "hidden" in order to view only the active site of an enzyme, for example. Promodeler I performs distance and angle measurements, and atoms may be labelled in all representations. The system accepts either Brookhaven Protein Data Bank structure files or user model files in the Brookhaven format. The software can be licensed for \$750 (US), or an entire hardware/software package consisting of Promodeler I software plus the Vermont Microsystems IM-640 can be purchased for \$2,000 (US).

The latest version of Autoscribe Ltd's molecular presentation graphics software includes a driver for Imagemaker, a device that **produces 35 mm slides** of chemical diagrams (Reader Service No. 109). Imagemaker is a desktop slide maker that Autoscribe says provides 8,000 line resolution and eliminates the pixels, stair-step discontinuities and jagged edges produced by the conventional CRT method. Imagemaker is sold as a package including the Imagemaker unit, Imagemate software and Autoscribe's molecular presentation graphics software for £4,000 (UK).

Computing on the edge

Perkin-Elmer Cetus, a joint venture between the two companies, has produced ELISAssoft — data analysis software to automate the **acquisition and manipulation of ELISA data** (Reader Service No. 110). The ELISAssoft package collects, evaluates, reports and stores data from the Perkin-Elmer Lambda Reader and a variety of other microplate readers. ELISAssoft performs regression analysis of the ELISA standard curve and a precision profile of the assay, and produces absorbance screening reports in graphical or tabular formats. Data can be presented in absorbance versus concentration, absorbance versus log, log versus log, or logit versus log forms. ELISAssoft is available in both IBM PC and Apple Macintosh versions, and costs \$950 (US).

Now researchers can have access to **Japanese journals translated into English**, within six months after publication. University Microfilms International and



Japanese journals give up their secrets with UMI's translation database.

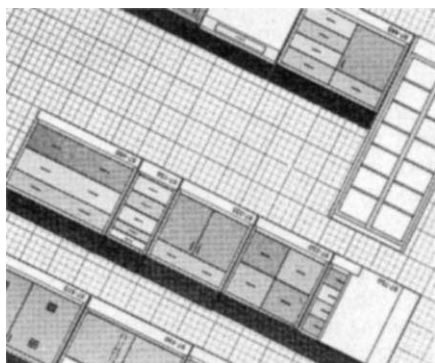
Dialog, one of the largest online databases, have collaborated to produce Japan Technology, an information service geared toward reversing the one-way flow of information exchange from West to East (Reader Service No. 111). UMI subscribes to nearly 600 Japanese journals, covering areas such as biotechnology, chemistry, fibre optics, ceramics, robot-

ics, production management, quality control and marketing. Selected articles are translated and abstracted, and entered into the database at the rate of 5,000 abstracts per month. The current database holds about 75,000 article abstracts, starting with 1985 issues. Prices run from \$90 (US) per connect hour for current subscribers to the print publication *Japanese Technical Abstracts*, to \$120 (US) per hour for non-subscribers. UMI says this means 10 key articles can be searched and printed out for \$25 to \$35 (US).

Laboratory setups

Lepco Incorporated has introduced a new generation of **clean rooms** that can deliver Class 100 contamination control with partially filtered ceilings (Reader Service No. 112). Lepco says the clean rooms provide true laminar clean air flow regardless of variations in HEPA filter medium or plenum pressure, and are unaffected by lights and blank-off panels. The membrane diffusion design also cuts down on turbulent air cones—a major problem with conventional clean rooms—according to the company. Lepco's grid is installed below HEPA filters and lights, and its porosity and translucence allows light and laminar air flow through its plastic surface. Suitable for control levels from Class 10 to Class 100,000, Lepco's system allows temperature and humidity to be controlled over the uniform plane of the room.

To help lab managers make the most out of limited space, Fisher Scientific offers a free **laboratory planning kit** (Reader Service No. 113). The kit comes

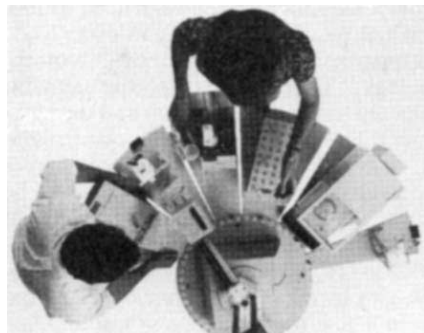


Where should the centrifuge go? Fisher Scientific helps you find the answer.

with over 150 different punch-out Fisher laboratory furniture pieces, including cabinets, wall cases, floor cases, fume hoods, desks, sinks, storage for flammables, tables, stools, undercounter refrigerators, filing systems, safety showers and eye-wash stations. The punch-outs are scaled 1:12, and can be arranged and rearranged on plasticized layout sheets until the optimum configuration is reached. When the perfect layout has been achieved, the pieces can be affixed to the sheet with peel-back adhesive. Fisher says its kit is useful in

planning a new installation, adding benchspace to an existing one, or merely visualizing ways to rearrange the furniture a laboratory already has. The kit comes complete with a list of tips for lab planning.

Zymark, one of the pioneers in **laboratory robotics**, has made automating laboratory procedures as easy as pie. Its new PyTechnology system, based around the



Zymark zaps the tedium of performing repetitive processes with the PyTechnology laboratory robotics system.

Zymark II robot, has pie-shaped task sections that can be added or taken away as needed to perform multiple-step procedures (Reader Service No. 114). The pre-configured, preprogrammed PySections contain all the hardware and software necessary to automate weighing, pipetting, dilution, filtration or centrifugation steps. Automated laboratory routines are implemented by writing procedures from a list of commands included with each PySection, and confirmation and error-detection are built into the software. Zymark says the PyTechnology system relieves technical staff from repetitive work, while also ensuring more precise data by eliminating operator-to-operator variations in results. An average system, consisting of the Zymark II robot and controller and 10 or 12 PySections, costs between \$45,000 (US) and \$55,000 (US).

A better mousetrap

It looks like a snake, but Research Products International's Zee-Arm is really meant to hold **pH electrodes, temperature probes, thermometers and other laboratory items** in place (Reader Service No. 115). Constructed of a polymer that is resistant to most laboratory chemicals, the Zee-Arm can be twisted or turned into any angle desired. It is mounted on a weighted base for stability, and is sturdy enough to hold a beaker over a bunsen burner. Research Products International charges \$99 (US) for the Zee-Arm.

Often, tissues and cultures that form the backbone of someone's research are exchanged through the mail. Polyfoam Packers Corporation has developed a **foam protective mailer** for tissue culture flasks to fill this need (Reader Service No.

116). The biomailers are available to fit a wide range of flask sizes. Mailers to fit 250 ml culture flasks cost \$34.90 (US) for a case of 24. The mailers have wall thicknesses of between 0.5 to 1.5 inches to protect flasks from breakage and vibration damage, and protect cultures from heat or cold. The mailers weigh less than corrugated mailing cartons so they are cheaper to send, and cardboard mailing sleeves are also available.

Setting it straight

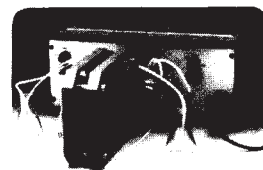
A REVIEW of the cDNA Synthesis System from Bethesda Research Laboratories, published in the 2 April issue of *Nature* (326, p. 526), contained several inaccuracies. BRL's system synthesizes cDNA using a cloned enzyme, Moloney murine leukaemia virus reverse transcriptase, and contains enough reagent for five reactions (with 2–10 µg of mRNA per reaction). □

These notes are prepared by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.

ADVERTISEMENTS

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5–10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM

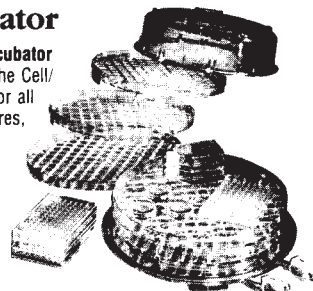
POB 126, D-8000 Munich 65, FRG

Reader Service No.49

Versatile Tissue Culture Incubator

Modular Incubator Chamber (the Cell/Sphere™) for all tissue cultures, aerobic or anaerobic.

Low Cost



Eliminates evaporation loss and contamination. Easy to use — flush with gas mixture, seal, place at appropriate temperature. Stackable, see-through. Quantity discounts. **billups-rothenberg, inc., pob 977, Del Mar, CA USA 92014. Call 619-755-3309 or contact your Flow Laboratories representative.**

Reader Service No.174

Biotechnology, where to next?

Richard Pearson

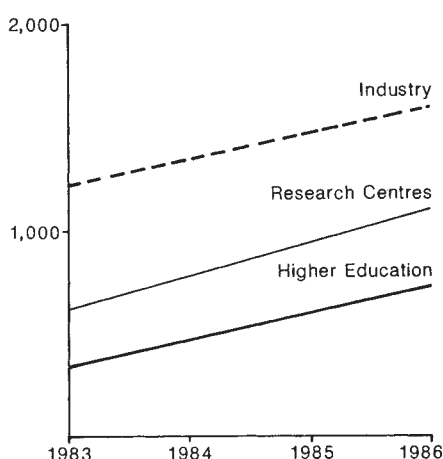
Facilities for postgraduate training in the United Kingdom must be maintained in order to meet the steady demand for postdocs in biotechnology.

IN the late 1970s and early 1980s biotechnology was the glamour technology attracting many millions of pounds from venture capitalists and governments alike. Since then some of the glamour has gone, as have some of the jobs, as several of the new venture capital companies have declared redundancies and others have been taken over by the established multinationals. What is clear is that the pay-back period for this technology is going to be considerably longer than for micro-electronics for instance, stretching into the next century. Initially, the Western world, and especially the United States, the United Kingdom and West Germany, was ahead of Japan in this new technology, building on its historically more developed basic scientific research. However, in the past five years Japan has caught up fast and now has a considerable amount of expertise in molecular biology. It has several large research groups augmented by large numbers of postgraduate research students, often as many as 25 per professor. Government funding of academic research is now substantial and large collaborative programmes between universities and industrial laboratories are thriving, often with extensive public financial support.

One of the keys to success in the new technologies is the availability of skilled personnel, and in the case of biotechnology much of this is at the postdoctoral level. In this respect the United Kingdom has been a world leader. The UK Science and Engineering Research Council (SERC) has a key role in biotechnology, supporting research and postgraduate training and encouraging joint ventures between industry and higher education. SERC has just published a report¹ monitoring trends in the supply and demand for key skills. Since 1983 it has been estimated that the total number of professional technical staff in biotechnology has grown by two-thirds to a total of about 3,500 in 1986. This increase has been largely due to expansion in research centres and higher education, primarily at postdoctoral level and involving staff on short-term contracts. The growth in joint ventures is increasingly blurring the dividing lines between industry, research institutes and higher education, as academic bodies set up industrial service companies, research centres become major sources of research students and postdoctoral training, and industry becomes actively involved in higher education. Nevertheless there are

still many sectoral differences.

In the industrial sector the number of companies has doubled since 1983 to 100, fuelled by new start-ups and established companies moving into novel biotechnology. The main emphasis remains heavily focused on research and development. However, production, plant engineering and industrial service companies are developing, although it is likely to be some years before they have significant employment levels. The majority of companies are, however, still small, half employing less than ten professional staff but with a few major companies accounting for half of the 1,600 in this sector. Among the 20 or so research centres employing



Changes in employment of personnel in biotechnology for the years 1983–1986 (from ref. 1).

about another 1,000 staff, there was a greater concentration of employment with five employing 100 or more such staff. Finally, there are more than 120 university, polytechnic and medical school departments active in novel biotechnology, employing a further 700 professionals — a doubling since 1983 and the fastest rate of growth in the three sectors (see graph).

Biotechnology is an interdisciplinary activity involving scientists and technologists in a range of bioscience and bioprocess skills that cut across conventional academic boundaries. Nowadays they can best be classified under three broad skill groups: genetic engineering (the fastest growing group), hybridoma technology, and bioprocess technology. In qualification terms the emphasis remains on PhD and postdoctoral levels, the only exception being the biochemical engineers, where industrial companies recruit a few specialist MSc graduates to work with large-scale reactors. First degree gradu-

ates have only been in demand as scientific assistants or higher-level technicians. Although employment has expanded significantly, recruitment to individual organizations has typically been measured in ones or twos each year, with only seven industrial firms and two research institutes recruiting ten or more such people in the past year. Recruitment difficulties for industry and research centres have been largely confined to highly specialist posts, usually at senior level, where the quality rather than the quantity of applicants was the problem.

The main recruitment problems have been experienced in higher education where the problems were being compounded by the vacancies only offering short-term contracts and relatively low salaries combined with generally poor facilities and low morale. The study found that the brain drain in biotechnology had in fact lessened and was now principally focused on younger scientists for whom few opportunities existed in higher education, and with the growth in demand in industry offering more opportunities in the UK for experienced people. The slowdown in growth and the maturing of biotechnology in North America has also been reducing the demand for migrants.

Looking ahead, there will be further growth in demand, although it is likely to be dominated by short-term posts in research centres and higher education. Selected shortages are likely to continue and become more widespread, particularly in relation to skills in plant molecular biology and in aspects of bioprocess technology. The most important factor is likely to be public expenditure, which on the one hand may boost demand but on the other may not keep pace with the growing need for postgraduate training, especially the need for increased levels of supervisory time and the associated equipment costs. At present, skill shortages primarily affect higher education, and if this remains the case then it will not be too long before they feed through and affect the research centres and industry alike and act as a serious constraint on the United Kingdom's development of biotechnology. □

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.

1. Bevan, S., Parsons, D. & Pearson, R. *Monitoring the Biotechnology Labour Market* (Science and Engineering Research Council, 1987).